

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 18:47:47 ; Search time 1988.65 Seconds
(without alignments)
8552.833 Million cell updates/sec

Title: US-09-665-728-2
Perfect score: 1031
Sequence: 1 tctagcgaaccttcggcc.....agccggctcgagcgccgc 1031

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb.ba:*

2: gb.htg:*

3: gb.in:*

4: gb.om:*

5: gb.ov:*

6: gb.pat:*

7: gb.ph:*

8: gb.pl:*

9: gb.pr:*

10: gb.ro:*

11: gb.sts:*

12: gb.sy:*

13: gb.un:*

14: gb.vi:*

15: gb.ba:*

16: em.fun:*

17: em.hum:*

18: em.in:*

19: em.om:*

20: em.or:*

21: em.ov:*

22: em.pat:*

23: em.ph:*

24: em.pl:*

25: em.ro:*

26: em.sts:*

27: em.sy:*

28: em.un:*

29: em.vi:*

30: em.htgo_hum:*

31: em.htgo_inv:*

32: em.htgo_rod:*

33: em.htg_hum:*

34: em.htg_inv:*

35: em.htg_rod:*

36: em.htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1031	100.0	1031	6	AX108634	AX108634 Sequence
2	336.8	32.7	18207	9	AF156100	AF156100 Homo sapi
3	305.8	29.7	191750	2	AC024113	AC024113 Mus muscu
c 4	215.8	20.9	191750	2	AC024113	AC024113 Mus muscu
5	215.8	20.9	230755	2	AC024618	AC024618 Mus muscu
6	185.2	18.0	15597	3	AF074901	AF074901 Caenorhab
7	183.4	17.8	182211	9	AL360004	AL360004 Human DNA
8	108.4	10.5	41345	3	CEP15G9	CEP15G9 Human DNA
9	93.6	9.1	1610	9	HS245418	HS245418 Homo sapi
10	91.4	8.9	4312	10	AF134319	AF134319 Mus muscu
11	89.4	8.7	157029	9	AL135796	AL135796 Human DNA
12	69.8	6.8	87461	9	HS211861	HS211861 Human DNA
c 13	69.8	6.8	180248	9	AC023275	AC023275 Homo sapi
c 14	65.4	6.3	717	11	HSJ348T1	HSJ348T1 Homo sapi
15	61.6	6.0	356	11	HSJ348T1	HSJ348T1 Homo sapi
16	51.6	5.0	1873	3	NCU37520	NCU37520 H.sapiens (
17	50.4	4.9	2336	3	NEPDSF	NEPDSF Nephila cla
18	50.4	4.9	2338	6	AR088543	AR088543 Sequence
19	50.4	4.9	2338	6	I92789	I92789 Sequence 1
20	50.2	4.9	4359	10	D85391	D85391 Mus musculu
21	48.6	4.7	87810	2	AC022987	AC022987 Homo sapi
22	47.6	4.6	154137	8	AP003215	AP003215 Oryza sat
c 23	47	4.6	41173	1	SCI41	SCI41 Streptomy
c 24	46.2	4.5	3867	1	AY033407	AY033407 Myxococcu
25	46	4.5	1026	3	AF350272	AF350272 Gasteraca
26	44.8	4.3	43349	1	SCR41	SCR41 Streptomy
27	44.6	4.3	4377	10	RNU62897	RNU62897 Rattus norv
c 28	44.2	4.3	44109	1	SCSF1	SCSF1 Streptomy
29	44	4.3	12606	1	AE005995	AE005995 Caulobact
c 30	43.8	4.2	18023	1	AX024392	AX024392 Sequence
c 31	43.8	4.2	18023	6	AX024285	AX024285 Sequence
c 32	43.8	4.2	87810	2	AC022987	AC022987 Homo sapi
33	43.4	4.2	2696	1	AF083621	AF083621 Lysobacte
34	43.4	4.2	2745	6	AX179756	AX179756 Sequence
35	43	4.2	2047	10	BC003909	BC003909 Mus muscu
36	43	4.2	5991	10	AF199422	AF199422 Mus muscu
37	43	4.2	6423	10	AF199421	AF199421 Mus muscu
c 38	43	4.2	198677	1	AE001863	AE001863 Deinococc
39	42.6	4.1	132470	9	HS203P18	HS203P18 Human DNA s
40	42.4	4.1	3203	10	AF008561	AF008561 Mus muscu
41	42.4	4.1	12356	10	AF134318	AF134318 Mus muscu
42	42.4	4.1	135545	10	MMHC213L3	MMHC213L3 Mus muscu
c 43	42.4	4.1	224020	2	AC087117	AC087117 Mus muscu
44	42.2	4.1	100000	9	AP000503	AP000503 Homo sapi
45	42.2	4.1	125350	2	AC020768	AC020768 Homo sapi

ALIGNMENTS

RESULT 1
AX108634
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AX108634
Sequence 2 from Patent WO0123419.
AX108634.1 GI:13923866

PAT

30-APR-2001

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa;

Mammalia; Eutheria;

Rattus.

1 (bases 1 to 1031)

Stanton, L.W. and Kapoun, A.M.

Differentially expressed genes

Patent: WO 0123419-A 2 05-APR-2001;

SCIOS INC. (US)

Location/Qualifiers

1..1031

/organism="Rattus norvegicus"

source

```

/db_xref="taxon:10116"
96. .923
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37763.1"
/db_xref="GI:13923867"
/translation="MTFPAQLLPILLATVTVAAVTSDEPTKTLSPATGDTALFVF
DTVGSMDLDLQVLDIGASRLILKAASVAVNPFSYIVFSDAKDYGHKNNELLQLQL
RELRLYVGGDGDCEPKSYGALKAFEEASTSSGVQLDQKQVSEVLKWKVESAIQASK
KQSVYFVLTDGCDRTHPGFKLAEETVLSLSPGPEIEVRDPLGMSQGSPLLMQD"
VHLLSADHEEGEHPWRIPFDSUKETVLSLSPGPEIEVRDPLGMSQGSPLLMQD"

BASE COUNT      215 a   317 c   308 g   191 t
ORIGIN
Query Match      100.0%; Score 1031; DB 6; Length 1031;
Best Local Similarity 100.0%; Pred. No. 1.5e-196;
Matches 1031; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tctagcaacccttcggccgctagagcagactcactgcactctatccctgcgact 60
Db 1 TCTAGGAACCCCTTCGGCCCGCTAGACGAGACTGCACCTCCCATCTATCCCTGCGACT 60

QY 61 ggcgtccattaggctgcagctccggctcagcatgacgcttagggcgagctcctgc 120
Db 61 GCGCGTCCATTAGGGCTGCAGCCTCCGGCTCAGCATGACGCTTAGGGCGAGCTCCTGC 120

QY 121 cgtgctcctggcgacctactacagtagtggcgccgctcacatctgtagagccacga 180
Db 121 CGTGTCTCTGGCGACCTACACAGTAGTGGCGGGCGGGTTCACATCTGATGAGCCACGA 180

QY 181 agacgtgtcccgccacaggagagcgcacccctggccttgccttcgatgtcacggct 240
Db 181 AGACGCTGTCCCCCGCCACAGGAGAGCGCACCCCTGGCCTTCGTCTTCGATGTCAACGGCT 240

QY 241 ccatgtggagcagatctgatcgaggtgatcgacggcgccctcagcattcttgagcgagtc 300
Db 241 CCATGTGGGACGATCTGTATCGAGGTGATCGACGGCGCCTCAGCATCTTGGAGCGCAGTC 300

QY 301 tgaacgcccagcggcggtcatcgccaaactatgcgtctggtgccttcacagaccagaca 360
Db 301 TGAGCAGCCGACGGCGGTTCATCGCAACTATTCGCGTGGTGCCTTTCCACGACCCAGACA 360

QY 361 ttggcccagtgacctcaacggcagaccagtggtgtttcagagagagctgagacaactct 420
Db 361 TTGGCCCCAGTGACCCCTCACGGCGGACCCAGTGGTGTTCAGAGAGAGCTGAGACAACCTCT 420

QY 421 atgttcaggagggtgtgtgactgccagaaatagatgtggggccatcaagctgcgtgg 480
Db 421 ATGTTTCAGGAGGTGTGTACTGCCAGAAATGAGTGTGGGGGCCATCAAGGTCGCGTGG 480

QY 481 aggttgcgaaccccgctccttcactcagctcttcctgcgatgcctgcgaaggactacc 540
Db 481 AGGTTGCCAACCCCGCTCCTTCATCTACGTCTTCCTCGGATGCCGTGCCGCAAGGACTACC 540

QY 541 acaagagaatgagctccttcagctccttcgagctgaagcagatgcagagtggtcttcgctg 600
Db 541 ACAAGAAGATGAGCTCCTGCAAGCTTCCTGCAAGTGAAGCAGTGCAGAGTGGTCTTCGTCG 600

QY 601 tgactgggagctcggtgacccacccaccctggctacctggcttttgaggagatgcgct 660
Db 601 TGACTGGGGACTCGGGTGACCCGCCACCCACCCCTGGCTACCTGGCTTTTGGAGGAGTGCCT 660

QY 661 ccacagttcttggccaagtgttccagctggacaacgacagcagtgatcgaggtgttaagt 720
Db 661 CCACCACTTCTGGCCAAGTGTTCACGCTGGCAACGACGAGGTGTGGAGGTGTAAAGT 720

QY 721 ggggtgagctccgcatccagggctccaagattcatctgtctcagcagacacagagagg 780
Db 721 GGGTGGAGTCCGGCATCCAGCCCTCCAAAGTTTCATCTGCTGTCAGCACACACAGGAGG 780

QY 781 agggcgaaacacatggagaaatccctttgaccccgcttgaggaagtccaccatctcac 840

```

[illegible]

AUTHORS

Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
 Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,
 Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
 Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
 Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
 Fernandez, C., Ferriguto, D., Forcum-Tansey, J., Gill, R.,
 Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hoques, M.,
 Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
 Kovar, C., Liu, J., Liu, W., Loulseged, H., Lozado, R.J., Martin, R.,
 Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
 Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Oguh, M., Parish, B.,
 Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
 Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodson, A.,
 Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
 Worley, K. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 191750)
 Worley, K.C.

Direct Submission
 Submitted (24-FEB-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jan 4, 2001 this sequence version replaced gi:11096415.

COMMENT

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: MABZ
 Center clone name: RP23-277F6
 ----- Summary Statistics -----
 Sequencing vector: M13; L08821
 Chemistry: Dye-primer Bodipy: 74% of reads
 Chemistry: Dye-terminator Big Dye: 26% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 179536 bases at least Q40
 Consensus quality: 173428 bases at least Q30
 Consensus quality: 179682 bases at least Q20
 Estimated insert size: 180449; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 3.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 29 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 19880: contig of 19880 bp in length
 * 19881 19980: gap of unknown length
 * 19981 38620: contig of 18640 bp in length
 * 38621 38720: gap of unknown length
 * 38721 51091: contig of 12371 bp in length
 * 51092 51191: gap of unknown length
 * 51192 63586: contig of 12395 bp in length
 * 63587 63686: gap of unknown length
 * 63687 76765: contig of 13079 bp in length
 * 76766 76865: gap of unknown length
 * 76866 92384: contig of 15519 bp in length
 * 92385 92484: gap of unknown length
 * 92485 105754: contig of 13270 bp in length
 * 105755 105854: gap of unknown length
 * 105855 114346: contig of 8492 bp in length
 * 114347 114446: gap of unknown length
 * 114447 122760: contig of 8314 bp in length
 * 122761 122860: gap of unknown length
 * 122861 130073: contig of 7213 bp in length

* 130074 130173: gap of unknown length
 * 130174 135344: contig of 5171 bp in length
 * 135345 135444: gap of unknown length
 * 135445 140753: contig of 5309 bp in length
 * 140754 140853: gap of unknown length
 * 140854 146305: contig of 5452 bp in length
 * 146306 146406: gap of unknown length
 * 146407 150873: contig of 4468 bp in length
 * 150874 150973: gap of unknown length
 * 150974 15410: contig of 4437 bp in length
 * 15411 15510: gap of unknown length
 * 15511 159684: contig of 4174 bp in length
 * 159685 159784: gap of unknown length
 * 159785 163180: contig of 3396 bp in length
 * 163181 163280: gap of unknown length
 * 163281 165459: contig of 2179 bp in length
 * 165460 165559: gap of unknown length
 * 165560 168279: contig of 2720 bp in length
 * 168280 168379: gap of unknown length
 * 168380 170847: contig of 2468 bp in length
 * 170848 170947: gap of unknown length
 * 170948 175115: contig of 4168 bp in length
 * 175116 175215: gap of unknown length
 * 175216 177870: contig of 2655 bp in length
 * 177871 177970: gap of unknown length
 * 177971 181412: contig of 3442 bp in length
 * 181413 181512: gap of unknown length
 * 181513 183055: contig of 1543 bp in length
 * 183056 183155: gap of unknown length
 * 183156 184937: contig of 1782 bp in length
 * 184938 185037: gap of unknown length
 * 185038 187608: contig of 2571 bp in length
 * 187609 187708: gap of unknown length
 * 187709 189220: contig of 1512 bp in length
 * 189221 189320: gap of unknown length
 * 189321 190396: contig of 1076 bp in length
 * 190397 190496: gap of unknown length
 * 190497 191750: contig of 1254 bp in length.

FEATURES
source

1. 191750
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="11"
 /clone="RP23-277F6"

BASE COUNT 47260 a 47982 c 47655 g 46018 t 2835 others
 ORIGIN

Query Match 29.7%; Score 305.8; DB 2; Length 191750;
 Best Local Similarity 92.3%; Pred. No. 4.8e-52;
 Matches 322; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 19 cccgtagagcagactgcactgccatctatccctgcgacctgcgctgccattagggct 78
 |||||
 Db 191108 CCCGCTAGCGGAGACTGCATCGCATCTCTCACTGCGACCTCGGGTCCCATCAGGGCT 191167
 |||||
 Qy 79 gcagctccggctcagctagcagctaggggcgcagctcctcctgcgctcctcctgcgacct 138
 |||||
 Db 191168 GCAGCCTCCGGCTCACCATGACGCCCTGGGCGCAGCTCTGCGCATGCTCGTGGGATCT 191227
 |||||
 Qy 139 acacagtagtgccggcgagctcacatctatgacccacgagacgctgtcccccgcga 198
 |||||
 Db 191228 CTACAGCAGTGGCGGGTGGTTCAGCTGTGACGCCGCCACGAGACGGTGTCCCCGCCA 191287
 |||||
 Qy 199 caggagacgccaccctggccttcgtcttcgatgtccacggctccatgtgggacatctga 258
 |||||
 Db 191288 CAGGAGACGCCACCCTAGCCTTCGTTTCGATGTTCACCGGCTCCATGTGGGATGACCTGA 191347
 |||||
 Qy 259 tgcaggtgatcagcggcctcagcattcttgagcgcagctctgagcagcgcgcgggg 318
 |||||
 Db 191348 TGCAGGTGATCAGCGGCCCTCGCGCATTTCTGGAGCGCAGCTGAGCAGCGACGCCGG 191407
 |||||
 Qy 319 tcatgcccaactatcgctgggtgctttccacgacccacgacattggccc 367

```

Db 191408 TCATCCGCAACATGCGTGGTCCCTTCCACGACCCAGTAGAGCCCC 191456
|||||
RESULT 4
AC024113/c
LOCUS
DEFINITION AC024113 191750 bp DNA HTG 07-JAN-2001
SEQUENCE, 29 unordered pieces.
AC024113
VERSION AC024113.12 GI:12025591
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 191750)
Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwuonu, G., Carroll, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, X., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hoques, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Loussegh, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Wozny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
Murray, K. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 191750)
Worley, K.C.
Direct Submission
Submitted (24-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 4, 2001 this sequence version replaced gi:11096415.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MAB2
Center clone name: RP23-277F6
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 74% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 159536 bases at least Q40
Consensus quality: 173428 bases at least Q30
Consensus quality: 179682 bases at least Q20
Estimated insert size: 180449; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 19880: contig of 19880 bp in length
* 19881 19980: gap of unknown length

```

```

19981 38620: contig of 18640 bp in length
38621 38720: gap of unknown length
38721 51091: contig of 12371 bp in length
51092 51191: gap of unknown length
51192 63586: contig of 12395 bp in length
63587 63686: gap of unknown length
63687 76765: contig of 13079 bp in length
76766 76865: gap of unknown length
76866 92384: contig of 15519 bp in length
92385 92484: gap of unknown length
92485 105754: contig of 13270 bp in length
105755 105854: gap of unknown length
105855 114346: contig of 8492 bp in length
114347 114446: gap of unknown length
114447 122760: contig of 8314 bp in length
122761 122860: gap of unknown length
122861 130073: contig of 7213 bp in length
130074 130173: gap of unknown length
130174 133344: contig of 5171 bp in length
133345 135444: gap of unknown length
135445 140753: contig of 5309 bp in length
140754 140853: gap of unknown length
140854 146305: contig of 5452 bp in length
146306 146405: gap of unknown length
146406 150873: contig of 4468 bp in length
150874 150973: gap of unknown length
150974 155410: contig of 4437 bp in length
155411 155510: gap of unknown length
155511 159684: contig of 4174 bp in length
159685 159784: gap of unknown length
159785 163180: contig of 3396 bp in length
163181 163280: gap of unknown length
163281 165459: contig of 2179 bp in length
165460 165559: gap of unknown length
165560 168279: contig of 2720 bp in length
168280 168379: gap of unknown length
168380 170847: contig of 2468 bp in length
170848 170947: gap of unknown length
170949 175115: contig of 4168 bp in length
175116 175215: gap of unknown length
175216 177870: contig of 2655 bp in length
177871 177970: gap of unknown length
177971 181412: contig of 3442 bp in length
181413 181512: gap of unknown length
181513 183055: contig of 1543 bp in length
183056 183155: gap of unknown length
183156 184937: contig of 1782 bp in length
184938 185037: gap of unknown length
185038 187608: contig of 2571 bp in length
187609 187708: gap of unknown length
187709 189220: contig of 1512 bp in length
189221 189320: gap of unknown length
189321 190396: contig of 1076 bp in length
190397 190496: gap of unknown length
190497 191750: contig of 1254 bp in length.

```

FEATURES

Location/Qualifiers

1..191750

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="11"

/clone="RP23-277F6"

BASE COUNT 47360 a 47982 c 47655 g 46018 t 2835 others
 ORIGIN

Query Match 20.9%; Score 215.8; DB 2; Length 191750;
 Best Local Similarity 84.8%; Pred. No. 4.9e-34;
 Matches 267; Conservative 0; Mismatches 42; Indels 6; Gaps 2;

QY 709 aggtgttaaaagtggtgagtcgcgcacatccagcgcctccaaagttcactctgtcagcag 768
 |||||
 Db 13539 AGGTGTTAAAGTGGGTGGAGTGCAGCAATCCAGGCTCCAAAGTTTCATCTGCTGTCAG 13480

```
QY 769 accacgagagagagggcgaacacacatgagagatccccctttgaccccgcttgaaggag 828
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 13479 ACCATGAGGAGGAGGCGCAGACAGTGGAGAAATCCCTTTTACGCCACGCTGAAGGAGG 13420

QY 829 tcaccatctcactgacgagggccagggcctgagatcgaagtccggaccactgggtatgt 888
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 13419 TCACCATTCGACTGAGCGGACGAGGCGCTTGAGATTGAAGTCGGGACCCACTGGGTATGT 13360

QY 889 --ccagggttcacctcctctctctgatgcaagactgagctggaag---gccagctgag 942
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 13359 GCCCAGGGCTCACCTCTCTTTGGTATACAGACTGAGCTGAGAGACAAAAATGAG 13300

QY 943 gcgatggaagggggcctgagagatgctcagcccaataaaatctgctccacacaaa 1002
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 13299 GCCATGGGAGGAGGAGGACTGAGGAGATGCTCAGTCAGTAAATGCTACCTCACAACA 13240

QY 1003 aaaaaaaagcccg 1017
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 13239 TGAGGACAGAGTCTG 13225

RESULT 5
AC024618 LOCUS
DEFINITION AC024618 230755 bp DNA HTG 23-SEP-2000
MUS musculus clone RP23-22C1, WORKING DRAFT SEQUENCE, 45 unordered
pieces.
ACCESSION AC024618
VERSION AC024618.2 GI:10280861
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 230755)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
JOURNAL Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
REFERENCE Boguslavsky,L., Boukhaltier,B., Brown,A., Burkett,G.,
AUTHORS Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 23, 2000 this sequence version replaced gi:17139711.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5311
```

```
Center clone name: 22_C_1
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 206138 bases at least Q40
Consensus quality: 218628 bases at least Q30
Consensus quality: 223560 bases at least Q20
Insert size: 226355; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2864: contig of 2864 bp in length
* 2865 2964: gap of 100 bp
* 2965 4029: contig of 1065 bp in length
* 4030 4129: gap of 100 bp
* 4130 5611: contig of 1482 bp in length
* 5612 5711: gap of 100 bp
* 5712 6711: contig of 1000 bp in length
* 6712 6811: gap of 100 bp
* 6812 8073: contig of 1262 bp in length
* 8074 8173: gap of 100 bp
* 8174 9200: contig of 1027 bp in length
* 9201 9300: gap of 100 bp
* 9301 10629: contig of 1329 bp in length
* 10630 10729: gap of 100 bp
* 10730 12013: contig of 1284 bp in length
* 12014 12113: gap of 100 bp
* 12114 14003: contig of 1892 bp in length
* 14006 14105: gap of 100 bp
* 14106 15760: contig of 1655 bp in length
* 15761 15860: gap of 100 bp
* 15861 18390: contig of 2530 bp in length
* 18391 18490: gap of 100 bp
* 18491 20892: contig of 2402 bp in length
* 20893 20992: gap of 100 bp
* 20993 23366: contig of 2374 bp in length
* 23367 23466: gap of 100 bp
* 23467 26126: contig of 2660 bp in length
* 26127 26226: gap of 100 bp
* 26227 29160: contig of 2934 bp in length
* 29161 29260: gap of 100 bp
* 29261 33742: contig of 4482 bp in length
* 33743 33842: gap of 100 bp
* 33843 37207: contig of 3365 bp in length
* 37208 37307: gap of 100 bp
* 37308 41498: contig of 4191 bp in length
* 41499 41598: gap of 100 bp
* 41599 45456: contig of 3858 bp in length
* 45457 45556: gap of 100 bp
* 45557 50797: contig of 5241 bp in length
* 50798 50897: gap of 100 bp
* 50898 54710: contig of 3813 bp in length
* 54711 54810: gap of 100 bp
* 54811 59379: contig of 4569 bp in length
* 59380 59479: gap of 100 bp
* 59480 63217: contig of 3738 bp in length
* 63218 63317: gap of 100 bp
* 63318 66324: contig of 3007 bp in length
* 66325 66424: gap of 100 bp
* 66425 70575: contig of 4151 bp in length
* 70576 70675: gap of 100 bp
* 70676 75597: contig of 4922 bp in length
* 75598 75697: gap of 100 bp
* 75698 80005: contig of 4308 bp in length
* 80006 80105: gap of 100 bp
*
```


IMPORTANT: This sequence is not the entire insert of clone Fl5G9. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 The true left end of clone Fl5G9 is at 1 in this sequence. The true right end of clone Fl5G9 is at 4082 in sequence 247070.
 The true left end of clone T0989 is at 41245 in this sequence. The true right end of clone C43C3 is at 8835 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence 247067.
 The end of this sequence (41245..41345) overlaps with the start of sequence 247070.
 For a graphical representation of this sequence and its analysis see:- <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Fl5G9>.

FEATURES

Location/Qualifiers

```
1..41345
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="Fl5G9"
join(2873..3089,3208..3273,3480..3823,3869..3947,4131..4264)
/gene="Fl5G9.1a"
join(2873..3089,3208..3273,3480..3823,3869..3947,4131..4264)
/gene="Fl5G9.1a"
/note="cDNA EST yk102e12.3 comes from this gene
cDNA EST yk102e12.5 comes from this gene
cDNA EST yk168h5.3 comes from this gene
cDNA EST yk168h5.5 comes from this gene
cDNA EST yk450c6.3 comes from this gene
cDNA EST yk450c6.5 comes from this gene
cDNA EST yk268g4.3 comes from this gene
cDNA EST yk268g4.5 comes from this gene
cDNA EST yk252c2.3 comes from this gene
cDNA EST yk252c2.5 comes from this gene
cDNA EST yk527d8.3 comes from this gene
cDNA EST yk558g5.3 comes from this gene
cDNA EST yk584g11.3 comes from this gene"
```

/codon_start=1

```
/protein_id="CAA87331.1"
/db_xref="GI:3875951"
/db_xref="SWISS-PROT:Q10033"
/translation="MSSRTSYTPSPSRKYGWDYSSYSSTSRDRDGTGSYRDYR
SSTYSTPRYSTYASKYSTDLGKSTDKGDDYSAAPIAKDEIEELSDYAKTPTPT
ETVDAPREVLSQISDFVEEDDDQTEVEEVKLTLASPEESEEDEEDVRKAE
LAAASTQIRESSPVSPKVESSAALFANDNGNETNRTSPPTVLNAKLLGGIPWP
KSSDVKKEGGDAEALPKRYSDLLIARNTGVVEESKTTDDAYKNEYGAGTNGVKVSTH
NFA"
```

```
join(3213..3273,3480..3823,3869..3947,4131..4264)
/gene="Fl5G9.1b"
join(3213..3273,3480..3823,3869..3947,4131..4264)
/gene="Fl5G9.1b"
```

```
/note="cDNA EST yk527d8.5 comes from this gene
cDNA EST yk589g5.5 comes from this gene
cDNA EST yk584g11.5 comes from this gene"
```

/codon_start=1

```
/protein_id="CAB54215.1"
/db_xref="GI:5824447"
/db_xref="SPTREMBL:O9U3J1"
/translation="MDYSAAPTAKDAIEELSDYAKTPTPTETVDAPREVLSQEI
SDFVEEDDDQTEVEEVKLTLASPEESEEDEEDVRKAEELAASTQIRESSPVSS
PKEVSSAALFANDNGNETNRTSPPTVLNAKLLGGIPWPKSSDVKKEGGDAEALP
KKRVSDLIARNTGVVEESKTTDDAYKNEYGAGTNGVKVSTHFA"
complement(join(6076..6174,6498..6562,6959..7029,7077..7256,7305..7378))
/gene="Fl5G9.3"
```

```
complement(join(6076..6174,6498..6562,6959..7029,7077..7256,7305..7378))
/gene="Fl5G9.3"
/codon_start=1
```

```
/protein_id="CAA87333.1"
/db_xref="GI:3875953"
/db_xref="SWISS-PROT:Q10035"
/translation="MTHQPTLNYYQLKKQKQYIFLKFNSQLFWYIMFSTNDNAKYS
KIDFFHLLFQYAVKQTIHQKQSGSDPWDEFFYRIYSADKRDIDNAQCHSAGI
KLNLNKLDQKLNFTDYDSDINKLTPYAMSLWLNNSFFPSYLNQAGKRYAPMTKSDS
SE"
complement(join(8996..9122,9175..9413,9464..9559))
/gene="Fl5G9.2"
complement(join(8996..9122,9175..9413,9464..9559))
/gene="Fl5G9.2"
/codon_start=1
/protein_id="CAA87332.1"
/db_xref="GI:3875952"
/db_xref="SWISS-PROT:Q10034"
/translation="MDFNSLFLNSKIEDSDNDYRNSDCLSERPKKTSKTRSIKWS
ESDLTKMSKYDVAETMRMRNLEEVSNYSYFILLRQVSCQMDRLNAGTENILETG
EVSDELSKKKTFLMNMPRTAVLKVMTDPSNGDVEGVYRSLRFGMWYLR"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6348..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4b"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6417..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4a"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6417..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4a"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6417..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4a"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6417..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4a"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6417..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4a"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6417..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4a"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6417..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4a"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6417..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4a"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6417..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4a"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,19183..19581,19759..19881,19935..20171,20970..21059,21111..21198,21257..21441,21488..21590,23087..23202,23255..23363,23492..23693,23780..23939,23988..24077,24544..25504,25673..25874,25934..26313,26565..26955,27233..27476,27750..28067,28340..28517,28564..28842,28890..29002,36561..36632,36681..36762,37047..37126,38336..38407,38452..38575,38622..38773,38829..38952,39009..39082,39698..39828,39875..39939,40081..40209,40328..41345,247070.1:102..304,247070.1:350..522,247070.1:679..802,247070.1:1753..1890,247070.1:1941..2081,247070.1:2213..2770,247070.1:2834..3265,247070.1:3313..3426,247070.1:3882..4289,247070.1:4333..4473,247070.1:4523..4666,247070.1:4712..4837,247070.1:4944..5392,247070.1:5571..6141,247070.1:6417..6539,247070.1:6589..6774,247070.1:7235..7444,247070.1:7490..7687,247070.1:7828..7926,247070.1:7972..8171)
/gene="Fl5G9.4a"
join(13472..13724,14222..14326,15572..16215,16297..16402,16485..16806,16959..17346,17399..17570,17654..18468,18624..18747,18910..19133,1918
```

247070.1:2834. 3265,247070.1:3313. 3426,
247070.1:3882. 4289,247070.1:4333. 4473,
247070.1:4523. 4666,247070.1:4712. 4837,
247070.1:4944. 5392,247070.1:5571. 6141,
247070.1:6348. 6539,247070.1:6589. 6774,
247070.1:7235. 7444,247070.1:7490. 7687,
247070.1:7828. 7926,247070.1:7972. 8171)
/gene="F15G9.4a"
/note="contains similarity to Pfam domain: PF00047
(Immunoglobulin domain), Score=1411.6, E-value=0, N=47
cDNA EST yk556f1.5 comes from this gene"
/codon_start=1
/protein_id="CAA87336.1"
/db_xref="GI:3875956"
/db_xref="SPTREMBL:O76518"
/translation="MGRSPSLVGLVGLLLATTCSSYNDKNDPTGKSLAFVDIT
GSFMDLVQVREGAAIKFTVMAQREKLIYNIIVFPDHPYLGEIINTDSTTFMKQL
SVYVMTGDCGPEKTLTGILKALQISLPFSYFTDARSKDHLEVDINTFOERKOS
MYEARERGTVSRNIPVDKHLSELTISLQDKDSDNLDIVLRDPGRVTKLYSKE
GSTIDLNKVLIRLKDPSGCVWNTNSRLKHTIRVFGHCAVDKYGFAASRLDRLEL
ARRPVLNQDTYLLINMTGLIPGTGVEIDLVDYHGHSLYKAVASPHRTNPNMYFAGP
FVPPKGLFFVRVQGYDEDNTEFMRIAPTATGIVSGVPGFRAFMSPIHQEFVGRDLNLS

Query Match 10.5%; Score 108.4; DB 3; Length 41345;
Best Local Similarity 55.6%; Pred. No. 2.3e-12;
Matches 208; Conservative 0; Mismatches 166; Indels 0; Gaps 0;
Qy 474 gccgtgaggttgccaaaccccggtctcttcacgtctcttcggtgacccgtgcgaag 533
Db 15583 GCACCTTCAAAATTCCTACCGTCCTCATTCATTATGTGTCACTGATGCCGATCAAAG 15642
Qy 534 gactaccacaagaatgaagctcgcagctcgcagctgaagcagtcgcagctgcagtc 593
Db 15643 GATTACCACTAGAGATGAAGTCTTAATACAAATTAAGAAACAAAGCTCAGTTGTT 15702
Qy 594 ttcgtgactgctgggagtcggtgacgcgcacccacccctgctacgtgctgtttgaggag 653
Db 15703 TTCGTGATGACTGGTACTGGTAAATAGAACACATCCAGGATCCGTACATATGAAAA 15762
Qy 654 atcgctccacagcttgcgcagctgctcagctggagcagcagcagctgcgaggtg 713
Db 15763 ATCGCGCTGCTATCTTTTGGACAAAGTTTCCATCTTGAAGAGAGTGTGTCAGTACCCTA 15822
Qy 714 ttaaagtggtgagtcgcagcagcagcctcccaagtcactctctctcagcagcagcac 773
Db 15823 TTAGATATGTCAGCATGCTGTCAACACAGAGAAAGTTCATTGATGATCAGGACGCT 15882
Qy 774 gagggaggggggaacacacatggagaatcccttttgaccacagcttgaaagaagtcacc 833
Db 15883 GAACGGGAGGAACAGTGTCTAGAAACATCCAGTAGAACAGCATCTCTCAGAGCTCACC 15942
Qy 834 atctcaactgagcgg 847
Db 15943 ATCTCATTTGTCAGG 15956

RESULT 9
LOCUS HSA245418
DEFINITION Homo sapiens mRNA for G7c protein (G7c gene located in the class
III region of the major histocompatibility complex).
ACCESSION AJ245418
VERSION AJ245418.1 GI:5701855
KEYWORDS G7c gene; G7c protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1610)
AUTHORS Albertella,M.R., Jones,H., Thomson,W., Olavesen,M.G. and
Campbell,R.D.
TITLE Localization of eight additional genes in the human major

histocompatibility complex, including the gene encoding the casein
kinase II beta subunit (CSNK2B)
Genomics 36 (2), 240-251 (1996)
96411681
MEDLINE
REFERENCE 2 (bases 1 to 1610)
Albertella,M.R.
JOURNAL Thesis (1997) University of Oxford, Department of Biochemistry, MRC
Immunochimistry Unit
REFERENCE 3 (bases 1 to 1610)
Noek,M., Albertella,M.R., van Kooij,M., Wixon,J., van Vugt,H., de
Groot,K. and Campbell,R.D.
G7c, a novel gene in the mouse and human major histocompatibility
complex class III region, possibly controlling lung tumor
susceptibility
Immunogenetics 51 (4-5), 383-386 (2000)
20260998
MEDLINE
REFERENCE 4 (bases 1 to 1610)
Aguado,B.
JOURNAL Direct Submission
TITLE Submitted (02-AUG-1999) Aguado B., HGMP Resource Centre, MRC,
Genome Campus, Hinxton, Cambridge, CB10 1SB, UNITED KINGDOM
FEATURES
Location/Qualifiers
1. .1610
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
/note="class III region of the major histocompatibility
complex"
1. .1610
/gene="G7c"
<1. .>1610
/gene="G7c"
/function="unknown"
/codon_start=3
/product="G7c protein"
/protein_id="CA52192.1"
/db_xref="GI:5701856"
/translation="LGRDPSRLDITPASSLSFVLTDTGSMGEINAAKIQARHLVE
QRGSPMEPVHYLVLPFHPDGFPVFTTSDPSFMOQLNEHALGGDPEMCLSAQ
LALLHTPLSDIFVFTDASPKDAFLTNQVSLTOERSCRVTLFVTEVSRVQGRARE
ILSLRFEPIKAVALSAGGEVIFTKDQHIRDVAALVIGESMAALVTLLDPPVYVPGP
LVFSDVLLQKITVRIHGDIVQAAQSLDLFLHFGIPMEDGPHGPIPLTQPVAGLQTL
PQTGWEIQVTAEDTPGVRSQAQSLDLFLHFGIPMEDGPHGPIPLTQPVAGLQTL
LVETVGLSRANPGDPQPHFSLVILRGVPEAGLQVLPLEVPPEPGLLAASLPTL
LSTPRPFSLELIGODAAAGRLHRAAPQSTVPVLLLELSPGSLAPGSKVPLSLRIA
SFGSQPDLRLTFVNPFSLSLTSLNSRAHLENECLGPAVAGGPRFSGPFRGDCDC
RGRSPSTPDSCLRLLLVSPQPRGTG"

BASE COUNT 290 a 525 c 463 g 332 t
ORIGIN
Query Match 9.1%; Score 93.6; DB 9; Length 1610;
Best Local Similarity 52.8%; Pred. No. 5.2e-09;
Matches 226; Conservative 0; Mismatches 199; Indels 3; Gaps 1;
Qy 183 acgctgtccccccgacagagagcgcacccctgctctgctgtatgtaccggtcc 242
Db 24 AGGCTGCTGGACATCACCCTCCAGCTTGTCTTGCTGGACACCGGCAGC 83
Qy 243 atgtggagcagctcgtatgacgagcgcgctcagcattctggagcagctgtg 302
Db 84 ATGGGTGAGGAGATCAACGCTGCCAAAATCCAGGTCGCCACCTTTGTGGAGC---AGGCC 140
Qy 303 agcagccagcagcgggtcatcgcacactatcgctgggtgccttccacagccagacatt 362
Db 141 AGAGGACGCCCCATGAGCCTGTCCACTATGTCCTGGTCCCTTTTCATGACCCAGGGTTC 200
Qy 363 gcccagtagcactcagcgcgagcagcagtggtgttcagagagagctgagacaactctat 422
Db 201 GGCCTGTCTTTACACACAGTCACCTCTGCAGCAGCTTCTGCAGCAGCTTATGAGATCCAT 260
Qy 423 gtccaggagggtgtgtgactgccccgaaatagtgctgtggggccatcaagcgtccgtgagag 482

```

Db      261 GCCTGGGGGTGGAGACGACCTGAGATGTGCTCTGACCCCTCGAGCTGGCCCTGCTG 320
Qy      483 gttgcaacccggctctctcatcagctcttcttcgagtcgcccgtgcccgaagactaccac 542
Db      321 CACACACCTCCACTCTCAGATATCTTGTCTTCACGGATGCCTCCCTCCCAAGGATGCTTT 380
Qy      543 aagaagaatgagctctcagctcctcgcagctcgcagctgaagcagtcgagtggtcttcgtg 602
Db      381 CTCACCAACAGGTGGAATCCCTGACTCAGGAGCGCGCTGCGGGTAAACATTCCTGGTG 440
Qy      603 actggggga 610
Db      441 ACTGAAGA 448

RESULT 10
LOCUS   AF134319      4312 bp      mRNA      ROD      11-MAY-2000
DEFINITION Mus musculus G7c protein (G7c) mRNA, G7c-d allele, complete cds.
ACCESSION AF134319
VERSION   AF134319.1 GI:7381104
KEYWORDS house mouse.
SOURCE   Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Snoek,M., Teuscher,C. and van Vugt,H.
TITLE Molecular analysis of the major MHC recombinational hot spot
located within the G7c gene of the murine class III region that is
involved in disease susceptibility
JOURNAL J. Immunol. 160 (1), 266-272 (1998)
MEDLINE 98211706
REFERENCE
AUTHORS Snoek,M., Albertella,M.R., van Kooij,M., Wixon,J., van Vugt,H., de
Groot,K. and Campbell,R.D.
TITLE G7c, a novel gene in the mouse and human major histocompatibility
complex class III region, possibly controlling lung tumor
susceptibility
JOURNAL Immunogenetics 51 (4-5), 383-386 (2000)
MEDLINE 20260998
REFERENCE
AUTHORS Snoek,M., Van Kooij,A., van Vugt,H. and de Groot,K.E.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1999) Molecular Genetics, the Netherlands Cancer
Institute, Plesmanlaan 121, Amsterdam 1066 CX, The Netherlands
FEATURES
source location/Qualifiers
1. 4312
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/chromosome="17"
/map="19.00 cM; between G7a and G7"
/tissue_type="brain"
1. 4312
/gene="G7c"
/feature="note=D1/H6556E-3; genomic structure of the G7c gene is
presented in GenBank Accession Number AF134318; similar to
the human sequence in GenBank Accession Number AJ245418"
/allele="d"
205..2880
/gene="G7c"
/codon_start=1
/product="G7c protein"
/protein_id="AAF61401.1"
/db_xref="GI:7381105"
/translation="MLPVEVPLSHLGPPIILLQLLLPPTSAFFPNWISLLAAGSVT
HQDITTEALNVTLVLFLEQHPGRPLHVEDYRGFTLLADDFIAFYFGPFSRRFR
AALGEVSRANAQDFLPAPKSNPDHLDFDAERLVQGRTRLVGALRETLVAARALEYTLA
QRLGAALHALQDFYSHSNWELGERQHPHLLWPRQLWSLAQVGDPTSCDCEGLSC
PGNLDSTLTSTYFGMHPPKPFKCGSHGFGDQSSQPPRGGINDKDSTSPFSPHK
LHLQAEVALLASIEAFSLRLSRGDKAFSLRLDITPASSLSFVLDTGSMGEINAA

```

```

KIQARRIVEORQSGSPMEPVFVILVFFHDPFGFVFVTTSDPDSFWQKLEIHALGGDE
PMCLSLALEALLHPTPLSDIFVTDASPKDALLNRVESLFRERRCRVTLFVLTEDPS
RTGRRREALSPLEPEVEATARASGEVITTKDQYIQDVAAIYVESWAGLVTLPLD
PTGVPGPCVPSVDSLLKQVTVRMHGDISSEFWIKSPAGVSGCPREGTGLGHTRRFG
QFVTWTDPTGTWELQVAAAGTPRVVQATSLDFLHFHGISVEDGPHPLPLT
QVAGLQTOLLVEVTGLTSROKLVGGQPOFQFHVLRVREPTQLGRVLESGDPPVVRG
LLAASLPTLLSVSPFSLLELVGGGGESLRRTPAPQCSVAPVLELSGPPDFLTPG
SKAPLSHIVSFSGPQDLDLRTSVNPSFSLTSLNLRARLINESAWGLRLVLPDPSAA
PDSVVMVTVAAGCAGSOVPPTHAFRLRLVLAQSSKQDLQDGPAAHSAAPVLPVPSFALL
PSTLVTOGRAGGGMAGKAWMGTVGVFLFLLGCTSW"
1295..1524
/misc_feature
/feature="G7c"
/feature="note="exons around recombinational hot spot; corresponds
to sequence presented in GenBank Accession Number
AF008561"
polya_signal
/feature="G7c"
4254..4259
BASE COUNT 830 a 1310 c 1166 g 1006 t
ORIGIN

Query Match      8.9%; Score 91.4; DB 10; Length 4312;
Best Local Similarity 52.4%; Pred. No. 1.1e-08;
Matches 225; Conservative 0; Mismatches 201; Indels 3; Gaps 1;

Qy 183 acgtgtccccccacagggagacgcaccctggccttcgtcttcgtatgcacggctcc 242
Db 1123 AGGCTGTGGACATCACCCACGCTTCCAGCTTGACGTTGTCTGGACACACAGGCAGT 1182
Qy 243 atgtgggagcatctgatgcaggtgatcgacggcgctcacgcattcttgagcgagtcgt 302
Db 1183 ATGGCGGAGGAATCAACGACGACCAAGATCCAGGCTCGCCGATTTGTGGAGCAGCTCAA 1242
Qy 303 agcagccgcagccgggtcatcgccaaactatcgctgggtgctcttcacagaccagacatt 362
Db 1243 GGCACCCCAT---GGAGCCTGTTTCTATATCTCTAGTGCCTTCCACAGCCACGCGT 1299
Qy 363 ggcacagtgacccctcacggcgagccacagtggtgttcagagagagagctgagacaactctat 422
Db 1300 GGCCTCGCTTTTACAAACAGCAGCGACCCGACAGCTTTTGGCAGAACTCAACGAGATCCAT 1359
Qy 423 gtcaaggagagtggtgactgccagaaaatagtgctggggggccatcaaggtgcgtggag 482
Db 1360 CCCTTGGGGGGTGGAGATGACCCAGAGATGTGCCTGTCTGCCCTGAGCTAGCCCTGTG 1419
Qy 483 gttgcaaccccgctctctcatcagctcttcgagtcgagtcgagtcgagtcgagtcgagtc 542
Db 1420 CACACCCCTCCCTCTCTGACATCTTTGCTTCTACTGATGCTCTACCCAGGATGCTCTT 1479
Qy 543 aagaagaatgagctctcagctcctcagctcagctgagcagctgagctgagctgagctgagct 602
Db 1480 CTTACCAACCGGTGGAATCCTGTACTCGGGAGAGCGCGCTCGAGGTCACATTTCTAGTA 1539
Qy 603 actgggggac 611
Db 1540 ACTGAAGAC 1548

RESULT 11
LOCUS   AL135796      157029 bp      DNA      PRI      27-SEP-2000
DEFINITION Human DNA sequence from clone GSI-164L12 on chromosome 1 Contains
an STS and GSSs, complete sequence.
ACCESSION AL135796
VERSION   AL135796.6 GI:7706886
KEYWORDS HTG.
SOURCE   Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 157029)
AUTHORS Donnelly,S.
TITLE Direct Submission

```

JOURNAL

Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

COMMENT

On May 4, 2000 this sequence version replaced gi:7634143. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chrl>

GSL-164L12 is from the library Genome_Systems_ReleaseI VECTOR: phelOBAC11

This sequence is the entire insert of clone GSL-164L12 The true left end of clone GSL-15303 is at 103097 in this sequence.

FEATURES

source	
repeat_region	/note="L1M1 repeat: matches 5180. .5793 of consensus" 13614. .13877
repeat_region	/note="AluJo repeat: matches 27. .287 of consensus" 13879. .13981
repeat_region	/note="L1M2 repeat: matches 2048. .2147 of consensus" 13982. .14292
repeat_region	/note="AluYa5 repeat: matches 1. .311 of consensus" 14293. .15642
repeat_region	/note="L1M2 repeat: matches 640. .2048 of consensus" 15608. .15818
repeat_region	/note="L1PA15-16 repeat: matches 146. .337 of consensus" 15828. .16373
repeat_region	/note="L1PBA repeat: matches -752. -.208 of consensus" 16464. .16704
repeat_region	/note="L1M2 repeat: matches -676. -.428 of consensus" 16497. .17042
misc_feature	/note="match: GSS: Em:AQ004976" 17721. .19291
repeat_region	/note="L2 repeat: matches 1051. .2708 of consensus" 19471. .19528
repeat_region	/note="29 copies 2 mer aa 81% conserved" complement(21071. .21467)
misc_feature	/note="match: GSS: Em:AQ761948" 21985. .22024
repeat_region	/note="20 copies 2 mer ct 77% conserved" 22384. .22611
repeat_region	/note="MIR repeat: matches 3. .262 of consensus" 22697. .22863
repeat_region	/note="AluJb repeat: matches 137. .302 of consensus" 23754. .23861
repeat_region	/note="27 copies 4 mer acat 61% conserved" 24319. .24381
repeat_region	/note="AluJb repeat: matches 84. .148 of consensus" 24382. .24670
repeat_region	/note="AluX repeat: matches 22. .309 of consensus" 24671. .24822
repeat_region	/note="AluJb repeat: matches 148. .289 of consensus" 26950. .26985
repeat_region	/note="18 copies 2 mer tg 97% conserved" 26952. .26987
repeat_region	/note="9 copies 4 mer tgtg 97% conserved" 27687. .27917
repeat_region	/note="AluJo repeat: matches 83. .304 of consensus" 29345. .29476
repeat_region	/note="L2 repeat: matches 2605. .2750 of consensus" 30393. .30828
misc_feature	/note="match: GSS: Em:AQ825928" 30990. .31275
repeat_region	/note="AluX repeat: matches 14. .299 of consensus" 32321. .32866
misc_feature	/note="match: GSS: Em:AQ418559" 33201. .33344
repeat_region	/note="L2 repeat: matches 2562. .2696 of consensus" 33363. .33795
repeat_region	/note="L2 repeat: matches 2250. .2690 of consensus" 35271. .35548
repeat_region	/note="AluSq repeat: matches 18. .297 of consensus" complement(37223. .37713)
misc_feature	/note="match: GSS: Em:AQ667379" 37319. .37339
misc_feature	/note="match: STS: Em:G37800" 38212. .38679
repeat_region	/note="L1M1 repeat: matches 5847. .6324 of consensus" 40470. .40593
repeat_region	/note="MIR repeat: matches 8. .140 of consensus" 41014. .41123
repeat_region	/note="MIR repeat: matches 20. .141 of consensus" 41395. .41618
repeat_region	/note="MIR repeat: matches 10. .256 of consensus" 43230. .43494
repeat_region	/note="AluY repeat: matches 39. .307 of consensus" 43683. .44142
repeat_region	/note="L1MEC repeat: matches 1211. .1677 of consensus"

```

repeat_region 44767..45295
repeat_region /note="L1 repeat: matches 2530. .3129 of consensus"
repeat_region 45333..45571
repeat_region /note="AluJb repeat: matches 3. .41 of consensus"
repeat_region 45604..46929
repeat_region /note="L1P7 repeat: matches 4758. .6124 of consensus"
repeat_region 46930..47199
repeat_region /note="AluJb repeat: matches 24. .297 of consensus"
repeat_region 47356..47823
repeat_region /note="L1 repeat: matches 3647. .4127 of consensus"
repeat_region 47918..48310
repeat_region /note="L1 repeat: matches 4257. .4654 of consensus"
repeat_region 48311..48614
repeat_region /note="AluX repeat: matches 1. .308 of consensus"
repeat_region 48615..48796
repeat_region /note="L1 repeat: matches 4654. .4834 of consensus"
misc_feature 49029..49516
repeat_region /note="match: GSS: Em:AQ280129"
repeat_region 49087..49262
repeat_region /note="L1P repeat: matches 5164. .5357 of consensus"
repeat_region 49407..49510
repeat_region /note="L2 repeat: matches 2579. .2696 of consensus"
repeat_region 50481..51060
repeat_region /note="MLTIF repeat: matches 1. .536 of consensus"
misc_feature complement(51651..52071)
misc_feature /note="match: GSS: Em:AQ636105"
repeat_region 52290..52512
repeat_region /note="match: GSS: Em:AQ797420"
repeat_region 52936..53296
repeat_region /note="L1P15 repeat: matches 5785. .6156 of consensus"
repeat_region 53303..53330
repeat_region /note="7 copies 4 mer gtgt 96% conserved"
repeat_region 53355..53457
repeat_region /note="L1P15 repeat: matches 5689. .5791 of consensus"
repeat_region 53680..53695
repeat_region /note="108 copies 2 mer tt 57% conserved"
misc_feature complement(54152..54622)
repeat_region /note="match: GSS: Em:AQ568896"
repeat_region 54157..54274
repeat_region /note="MER45B repeat: matches 199. .318 of consensus"
repeat_region 54340..54856

```

```

Query Match 8.7%; Score 89.4; DB 9; Length 157029;
Best Local Similarity 64.0%; Pred. No.1e-08;
Matches 135; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 709 aggtgttaactgagtgagtcgccatccagcctccaaagttcattctgtcgcagcag 768
Db 125632 AGGTATTAAATGGTGAAGAGCAGTACAGGCTCCAAAGTTACCTTTTATCCACAG 125691

Qy 769 accagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 828
Db 125692 ATCATTTGGAACAGCCTGTAATACTTGGAGAAATCTTTTGATCCCGCTGAAAGAGG 125751

Qy 829 tcaccatctactagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 888
Db 125752 TCACTGTGCTTTGAGTGGGCTTCTCCAATGATGAATTCGCAATCTTTAGTGCAGA 125811

Qy 889 cccaggggttaccctctctctctctgacaga 919
Db 125812 TATATCAACATCATCAATAATAATACAAA 125842

```

```

RESULT 12
LOCUS HSG118G19 87461 bp DNA PRI 15-MAY-2001
DEFINITION Human DNA sequence from clone GSI-118G19 on chromosome 1q25.1-31.1, complete sequence.
ACCESSION AL121996
VERSION AL121996.15 GI:14140106
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 87461)
Thomas,D.
Direct Submission
Submitted (15-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 17, 2001 this sequence version replaced gi:13990347.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one W13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
GSI-118G19 is from the library Genome_Systems_ReleaseI VECTOR: pBelOBAC11
IMPORTANT: This sequence is not the entire insert of clone GSI-118G19. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone GSI-164112 is at 87362 in this sequence. The true right end of clone GSI-204112 is at 100 in this sequence.
Location/Qualifiers
1..87461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q25.1-31.1"
/clone="GSI-118G19"
/clone_lib="Genome_Systems_ReleaseI"
5..571
/note="L2 repeat: matches 2120. .2748 of consensus"
568..694
/note="L2 repeat: matches 2572. .2706 of consensus"
896..1200
/note="AluJb repeat: matches 1. .300 of consensus"
1292..1578
/note="AluJb repeat: matches 19. .297 of consensus"
1618..2109
/note="L1MA9 repeat: matches 5783. .6270 of consensus"
2110..2157
/note="24 copies 2 mer gt 95% conserved"
2166..2326
/note="AluSg/x repeat: matches 149. .309 of consensus"
2329..2936
/note="L1MA9 repeat: matches 5177. .5779 of consensus"
2936..3566
/note="L1MA9 repeat: matches 4500. .5146 of consensus"
3794..3891
/note="MER5A repeat: matches 13. .112 of consensus"
5634..5814
/note="MIR repeat: matches 59. .247 of consensus"
complement(5905..5949)
/note="Sequence confirmed by AC023275 sequenced by WUGSC"
6728..6926
/note="MIR repeat: matches 56. .259 of consensus"
7074..7136

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

```
repeat_region /note="MIR repeat: matches 50. .112 of consensus"
8362. .8848
repeat_region /note="LTR37B repeat: matches 6. .459 of consensus"
10408. .10856
repeat_region /note="MLTIC repeat: matches 2. .466 of consensus"
10884. .11020
repeat_region /note="MLT2B repeat: matches 210. .363 of consensus"
11056. .13925
repeat_region /note="L1PA3 repeat: matches 13. .2882 of consensus"
13926. .13962
repeat_region /note="Charlie3 repeat: matches 878. .914 of consensus"
13963. .17229
misc_feature /note="L1PA3 repeat: matches 2881. .6146 of consensus"
17162. .17277
misc_feature /note="Sequence confirmed by AC023275 sequenced by WUGSC"
17278. .17310
misc_feature /note="Sequence from AC023275 sequenced by WUGSC."
17311. .17418
repeat_region /note="Sequence confirmed by AC023275 sequenced by WUGSC"
17930. .18002
repeat_region /note="L1P4 repeat: matches 5695. .5768 of consensus"
18310. .18794
repeat_region /note="MLT1H repeat: matches 51. .543 of consensus"
20658. .20691
repeat_region /note="17 copies 2 mer tg 94% conserved"
20801. .21109
repeat_region /note="AluJb repeat: matches 1. .312 of consensus"
23465. .23676
repeat_region /note="MIR repeat: matches 2. .258 of consensus"
23750. .23852
repeat_region /note="L2 repeat: matches 2596. .2710 of consensus"
26376. .26824
repeat_region /note="MER65A repeat: matches 1. .445 of consensus"
26984. .27214
repeat_region /note="L1MC3 repeat: matches 7496. .7733 of consensus"
27243. .27862
repeat_region /note="L1MC3 repeat: matches 6695. .7336 of consensus"
27895. .27991
repeat_region /note="L1MC4 repeat: matches 6613. .6709 of consensus"
28061. .28121
repeat_region /note="L1MD3 repeat: matches 6536. .6597 of consensus"
29391. .29470
repeat_region /note="FLAWA repeat: matches 42. .121 of consensus"
29472. .29787
repeat_region /note="158 copies 2 mer at 78% conserved"
30307. .30453
repeat_region /note="MIR repeat: matches 79. .213 of consensus"
30455. .30751
repeat_region /note="Alusx repeat: matches 1. .297 of consensus"
30775. .30836
repeat_region /note="MIR repeat: matches 196. .260 of consensus"
30908. .30959
repeat_region /note="MIR repeat: matches 82. .140 of consensus"
32816. .32938
repeat_region /note="MIR repeat: matches 17. .142 of consensus"
33149. .33494
repeat_region /note="L2 repeat: matches 2355. .2710 of consensus"
34283. .35033
repeat_region /note="L1MEC repeat: matches 783. .1570 of consensus"
35069. .36448
repeat_region /note="L1MEC repeat: matches 1635. .2727 of consensus"
36454. .36607
repeat_region /note="MIR repeat: matches 8. .171 of consensus"
40801. .41093
repeat_region /note="AluJb repeat: matches 15. .312 of consensus"
41795. .42230
repeat_region /note="L1MB8 repeat: matches 5745. .6173 of consensus"
42231. .42396
repeat_region /note="L1M4 repeat: matches 5483. .5646 of consensus"
42531. .42960
repeat_region /note="L1PA16 repeat: matches 5721. .6157 of consensus"
43280. .44137
repeat_region /note="L2 repeat: matches 1880. .2750 of consensus"
```

```
repeat_region 44142. .44298
/note="MSTD repeat: matches 1. .134 of consensus"
repeat_region 44374. .44680
/note="Alusx repeat: matches 5. .311 of consensus"
repeat_region 44684. .44934
/note="MSTD repeat: matches 178. .426 of consensus"
repeat_region 44935. .45514
/note="L2 repeat: matches 1247. .1880 of consensus"
repeat_region 46831. .46872
/note="L2 repeat: matches 2709. .2750 of consensus"
repeat_region 47636. .47950
/note="AluY repeat: matches 1. .310 of consensus"
repeat_region 48164. .48425
/note="L2 repeat: matches 2431. .2710 of consensus"
repeat_region 48443. .48632
/note="L2 repeat: matches 2027. .2233 of consensus"
repeat_region 49480. .49851
/note="L2 repeat: matches 1235. .1662 of consensus"
repeat_region 49871. .49964
/note="MIR repeat: matches 20. .127 of consensus"
repeat_region 49991. .50252
/note="MLTIC repeat: matches 173. .456 of consensus"
repeat_region 50253. .50564
/note="Alusx repeat: matches 1. .312 of consensus"
repeat_region 50565. .50743
/note="MLTIC repeat: matches 1. .173 of consensus"
repeat_region 50981. .51277
/note="Alusx repeat: matches 1. .296 of consensus"
repeat_region 51693. .51718
/note="13 copies 2 mer tt 100% conserved"
repeat_region 52326. .52507
/note="MIR repeat: matches 76. .244 of consensus"
repeat_region 52558. .52629
/note="MIR repeat: matches 75. .149 of consensus"
repeat_region 52640. .52807
/note="MER63A repeat: matches 6. .174 of consensus"
repeat_region 53094. .53380
/note="L2 repeat: matches 1729. .2035 of consensus"
repeat_region 53456. .53885
/note="L2 repeat: matches 2285. .2730 of consensus"
repeat_region 54316. .54481
/note="83 copies 2 mer tt 56% conserved"
repeat_region 55309. .55565
/note="L1MD repeat: matches 1211. .1474 of consensus"
repeat_region 56353. .56911
/note="L1M4 repeat: matches 3491. .4067 of consensus"

Query Match 6.8%; Score 69.8; DB 9; Length 87461;
Best Local Similarity 63.3%; Pred. No. 0.0001;
Matches 107; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 196 ccacaggagacgcacacctgcttcgtatgctcaccggtccatgtgggaagatc 255
|| || || || || || || || || || || || || || || || || || || ||
Db 38544 CCGAGGGGGCTCCACCTTGGCTTTTGTGTTGACTGTCTTATGATGATGATT 38603

QY 256 tgatgcaggtgatcgagcgccctcacgcattctgaagcagcttgagcagcgagcc 315
|| || || || || || || || || || || || || || || || || || || ||
Db 38604 TAGTTCAGGTGATTGAGGGGGCTCCAAAATTTGGAGACGCTTTGAAAAGACCTAAAA 38663

QY 316 gggatcatcgcaactatgcctgctgcttcctccacgaccagacattgg 364
|| || || || || || || || || || || || || || || || || || || ||
Db 38664 GACCTCTTTTCAACTTTGGCTTGGTGGCTTTCCATGATCCAGGTAAAGG 38712

RESULT 13
AC023275/c
LOCUS AC023275 180248 bp DNA PRI 09-MAY-2001
DEFINITION Homo sapiens BAC clone RP11-375110 from 1, complete sequence.
ACCESSION AC023275
VERSION AC023275.3 GI:9838169
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 180248)

Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

2 (bases 1 to 180248)

Swearengen, S., Mauplin, R., Drone, K. and Gregory, S.

The sequence of Homo sapiens BAC clone RP11-375L10

Unpublished

3 (bases 1 to 180248)

Waterston, R.H.

Direct Submission

Submitted (10-FEB-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 180248)

Waterston, R.H.

Direct Submission

Submitted (17-AUG-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 180248)

Waterston, R.H.

Direct Submission

Submitted (10-SEP-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

6 (bases 1 to 180248)

Waterston, R.H.

Direct Submission

Submitted (08-NOV-2000) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

7 (bases 1 to 180248)

Waterston, R.H.

Direct Submission

Submitted (09-MAY-2001) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Aug 17, 2000 this sequence version replaced gi:7630968.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@watson.wustl.edu

----- Summary Statistics

----- Center project name: H_NH0375L10

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.

McPherson, Department of Genetics, Washington University, St. Louis

MO. For additional information about the map position of this

sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male

donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,

Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved

approach for construction of bacterial artificial chromosome

libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-375L10;

Actual end is at base position 180248 of RP11-375L10.

FEATURES

Location/Qualifiers

1..180248

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/map="1"

/clone="RP11-375L10"

/clone_lib="RPCI-11"

667..820

/rpt_family="CRI"

926..1011

/rpt_family="CRI"

2272..2420

/rpt_family="MIR"

5140..5262

/rpt_family="L2"

6076..6593

/rpt_family="L2"

6767..7258

/rpt_family="L2"

7348..7653

/rpt_family="Alu"

7848..8645

/rpt_family="L2"

8938..9011

/rpt_family="MER1_type"

9685..9725

/rpt_family="L1"

9801..10110

/rpt_family="Alu"

12246..12330

/rpt_family="MIR"

12577..12788

/rpt_family="Alu"

12790..12957

/rpt_family="Alu"

13055..13209

/rpt_family="L2"

16393..17278

/rpt_family="AcHobo"

17344..17461

/rpt_family="AcHobo"

18161..18263

/rpt_family="L1"

18322..18687

/rpt_family="L1"

20558..21137

/rpt_family="MaLR"

22114..22321

/rpt_family="L2"

22353..22811

/rpt_family="L1"

22832..23003

/rpt_family="L1"

23004..23307

/rpt_family="Alu"

23308..23700

/rpt_family="L1"

23740..24219

/rpt_family="L1"

24419..24688

/rpt_family="Alu"

24689..26014

/rpt_family="L1"

26047..26106

repeat_region

```

/rpt_family="Alu"
26433..26642
/rpt_family="L1"
26654..26881
/rpt_family="L1"
27044..27199
/rpt_family="L1"
27270..27726
/rpt_family="L1"
28124..28388
/rpt_family="Alu"
29167..29275
/rpt_family="MIR"
30000..30223
/rpt_family="MIR"
30495..30677
/rpt_family="MIR"
31025..31217
/rpt_family="MIR"
32939..33406
/rpt_family="L1"
33434..33836
/notes="similar to
34177..34516
/notes="similar to
35831..35967
/rpt_family="CR1"
36070..36347
/rpt_family="Alu"
36643..36920
/rpt_family="CR1"
36969..37038
/rpt_family="CR1"
37823..38255
/rpt_family="L2"
38274..38387
/rpt_family="L2"
39802..39869
/rpt_family="L2"
40343..40628
/rpt_family="Alu"
43701..43931
/rpt_family="Alu"
46796..46947
/rpt_family="Alu"
46948..47236
/rpt_family="Alu"

Query Match      6.8%; Score 69.8; DB 9; Length 180248;
Best Local Similarity 63.3%; Pred. No. 8.3e-05;
Matches 107; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 196 ccacaggagagccaccctggccttcgtatgtcaccggtccatgtggagcagtc 255
|| || || || || || || || || || || || || || || || || || || ||
Db 120435 CCGAGGGGGCTCCACGCTGGCTTTGGTTGATGTGACGTGTTCTATGATGATT 120376

Qy 256 tgatcgagtgatcgagccgccacgcattctggagcgcagtcgtgagagccgagcc 315
|| || || || || || || || || || || || || || || || || || || ||
Db 120375 TAGTTCAGTGATGAGAGGGGCTCCAAAATTTGGAGACGCTTTGAAAAGACCTAAAA 120316

Qy 316 gggctacgcgaactatcgctgggtgcctttccacgaccagacagattgg 364
|| || || || || || || || || || || || || || || || || || || ||
Db 120315 GACCTCTTTCAACTTTCGTTGGTGCTTTCCATGATCCAGGTAAGGG 120267

RESULT 14
LOCUS HSJ34G10 717 bp DNA STS 07-MAR-2000
DEFINITION STS from H.sapiens random shear fragment, sequence tagged site.
ACCESSION AL159308
VERSION AL159308.1 GI:7210622
KEYWORDS STS.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 717)
AUTHORS Hunt,S., Sims,S., Willey,D., Carter,N. and Ross,M.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2000) E-mail contact: humquery@sanger.ac.uk
COMMENT Marker stSG68675 (Primer A : CTCCTCCAGATGGGACC;
Primer B : TGGATGTTATATCGTCTCG; amplicon size : 121 bp) was
developed from a single pass sequencing read from H.sapiens
flow-sorted chromosome 9-12 random shear fragment, SC9-12pJ34g10.
Vector : pUC18 Site : SmaI
Further information : http://www.sanger.ac.uk/HGP/Chr9/.
FEATURES
Location/Qualifiers
SOURCE 1..717
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/sex="male"
/clone="SC9-12pJ34g10"
/tissue_type="Lymphoblastoid cell line"
/clone_lib="SC9-12pJ"
BASE COUNT 186 a 198 c 211 g 120 t 2 others
ORIGIN
Query Match 6.3%; Score 65.4; DB 11; Length 717;
Best Local Similarity 72.4%; Pred. No. 0.0028;
Matches 84; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
Qy 330 tatcgctgggtgccttcacgaccagacattgcccagtcacctcacctcagcgagccca 389
|| || || || || || || || || || || || || || || || || || || ||
Db 557 TCTGGGGCTCCCTCTCTCTCCACAGATATTGGCNCAGTACCCCTCAGCGGACCC 498
Qy 390 gtgtgttcagagagagctgagacaactctatgttcaggaggtggagctgccc 445
|| || || || || || || || || || || || || || || || || || || ||
Db 497 ACAGTGTTCAGAGGAGCTGAGAGAACTCTACGTGCAGTGGGCGAGCCCTGACC 442

-RESULT 15
HSJ348TG1 356 bp DNA STS 22-MAR-1996
LOCUS H.sapiens (DIS2848) DNA segment containing (CA) repeat; clone
DEFINITION AFM348tg1; single read, sequence tagged site.
ACCESSION Z51502
VERSION Z51502.1 GI:1232802
KEYWORDS STS; CA repeat; dinucleotide repeat; GT repeat; microsatellite DNA;
microsatellite marker; repeat polymorphism.
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 356)
AUTHORS Wellssenbach,J.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1995) Genethon, B.P. 60, 91002 Evry Cedex France.
E-mail: Jean.Weissenbach@genethon.fr
Db (bases 1 to 356)
Dib,C., Faure,S., Fizames,C., Samson,D., Drouot,N., Vignal,A.,
Mollasseau,P., Marc,S., Hazan,J., Seboun,E., Lathrop,M., Gyapay,G.,
Morissette,J. and Wellssenbach,J.
COMMENT A comprehensive genetic map of the human genome based on 5,264
microsatellites
Nature 380 (6570), 152-154 (1996)
96176476
full automatic.
Location/Qualifiers
SOURCE 1..356
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/cell_line="CEPH 134702"
/clone_lib="genomic DNA"

```

BASE COUNT 111 a /note="cloning vector is M13mpl8"
ORIGIN 79 c 46 g 107 t 13 others

Query Match 6.0%; Score 61.6; DB 11; Length 356;
Best Local Similarity 68.7%; Pred.No. 0.02; Mismatches 36; Indels 0; Gaps 0;
Matches 79; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 580 agtcgagggtggtcttcgtgactggggactgagcgacccacccacccctggctacc 639
Db 242 AATCATAGGTCGTATTGTTCTGCTGNGATTGTGATCAGAGNCCCATATNGGTATA 301
QY 640 tggtctttgaggagatgcctccaccagttcttgccaaagtgtccagctggacaa 694
Db 302 AAGTCTATGGNGAAATGCCTCTNCAAGTTCTGTTCAAGTGTTCATCTGGACAA 356

Search completed: March 7, 2002, 20:44:19
Job time: 6992 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 18:53:02 ; Search time 1634.39 Seconds
(without alignments)
6778.613 Million cell updates/sec

Title: US-09-665-728-2

Perfect score: 1031

Sequence: 1 tctagcgaaccccttcggcc.....agccggctcgagcgccgc 1031

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estfun:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estom:*
- 5: em_estpl:*
- 6: em_estba:*
- 7: em_estro:*
- 8: em_estov:*
- 9: em_htc:*
- 10: gb_estl:*
- 11: gb_est2:*
- 12: gb_htc:*
- 13: gb_gss:*
- 14: em_gss_fun:*
- 15: em_gss_hum:*
- 16: em_gss_inv:*
- 17: em_gss_pin:*
- 18: em_gss_pro:*
- 19: em_gss_rod:*
- 20: em_gss_vrt:*
- 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	309	30.0	470	AI449932	mr81b03.x
C 2	90.4	8.8	720	AZ985073	2M0266L17
C 3	77.4	7.5	937	CNS025DE	AL181931 Tetraodon
C 4	70.6	6.8	559	BF418902	UI-R-BJ2-
C 5	64	6.2	915	AZ137837	SP-0178-A
C 6	61	5.9	528	BE508854	dcl3h07.y
C 7	55	5.3	754	AZ189426	SP-1014_A
C 8	49.6	4.8	935	CNS006AK	AL066051 Drosophil
C 9	48.8	4.7	925	CNS0091P	AL053013 Drosophil
C 10	46.6	4.5	520	BG895162	358378 MA
C 11	46.4	4.5	925	CNS0091P	AL053013 Drosophil
C 12	46	4.5	965	CNS02215	AL178322 Tetraodon

C	13	45	4.4	818	13	CNS03902	AL234371 Tetraodon
	14	44.8	4.3	375	11	BF481079	BF481079 Fm1_16_B0
	15	44.8	4.3	379	11	BG560654	PHIZ2_58
	16	44.8	4.3	392	11	BF481089	BF481089 Fm1_16_CO
	17	44.8	4.3	406	11	BG560698	BG560698 RHIZ2_58
	18	43.2	4.2	523	11	BG240204	OVI_19_B0
	19	43.2	4.2	924	13	CNS04SRD	AL305554 Tetraodon
C	20	42.6	4.1	543	10	AW124542	UI-M-BH2.
	21	42.2	4.1	461	10	AU174537	AU174537
	22	42	4.1	632	10	BG744240	602723347
C	23	41.8	4.1	503	10	AI653725	AI653725 WD36F06.x
	24	41.8	4.1	686	11	BG835990	BG835990 Zm06_06h0
C	25	41.4	4.0	384	10	AA294693	SMOV3MCA9
	26	41.4	4.0	476	10	BE477180	BE477180 160761 BA
C	27	41.4	4.0	566	10	AW082912	AW082912 xc04c11.x
C	28	41.4	4.0	872	11	BG345072	BG345072 HVSMBG001
	29	41	4.0	647	11	BG180123	BG180123 602329707
C	30	41	4.0	778	10	AI991270	AI991270 WU41h01.x
C	31	41	4.0	814	10	AW081400	AW081400 xc41h07.x
C	32	40.8	4.0	523	10	AI855212	AI855212 603010C06
C	33	40.8	4.0	556	11	BG874151	BG874151 MEST46-H0
C	34	40.8	4.0	584	13	AQ848793	AQ848793 LMAJFV1_1
C	35	40.8	4.0	586	10	BES53066	BES53066 946088H08
C	36	40.8	4.0	603	10	AI855218	AI855218 603010D06
C	37	40.8	4.0	626	10	AI855245	AI855245 603010H10
C	38	40.8	4.0	671	13	AQ901789	AQ901789 LMAJFV1_1
	39	40.8	4.0	813	11	BG321263	BG321263 Zm04_0390
C	40	40.8	4.0	956	13	CNS03GMI	AL243171 Tetraodon
C	41	40.6	3.9	460	10	AI432376	AI432376 tg54h07.x
C	42	40.6	3.9	573	10	BE452815	BE452815 894066A05
C	43	40.6	3.9	738	11	BG864336	BG864336 602798212
	44	40.4	3.9	459	10	AW504911	AW504911 UI-HF-BN0
	45	40.4	3.9	470	11	BF313055	BF313055 601896455

ALIGNMENTS

RESULT	1	AI449932/c	470 bp	mrna	EST	09-MAR-1999
LOCUS		mr81b03.x1	Stratagene mouse heart (#937316)	Mus musculus	cDNA clone	
DEFINITION		IMAGE:603821.3'	similar to WP:FI5G9.4	CE01552	IG SUPERFAMILY	
ACCESSION		AI449932	GI:4293462			
VERSION		EST.				
KEYWORDS		house mouse.				
SOURCE		Mus musculus				
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE		1 (bases 1 to 470)				
AUTHORS		Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, K., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.				
TITLE		The WashU-NCI Mouse EST Project 1999				
JOURNAL		Unpublished (1999)				
COMMENT		Contact: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. This clone was previously sequenced on the 5' end only, this new data is from the 3' end Possible reversed clone: polyT not found High quality sequence stop: 467.				
FEATURES		Location/Qualifiers				
		1..470				

```
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:603821"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
EORI: Site_2: XhoI; Vector: pBluescript SK-; Site_1:
Oligo dT, 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5',
adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTT 3'
80 a 152 c 168 g 69 t 1 others
BASE COUNT
ORIGIN
Query Match 30.0%; Score 309; DB 10; Length 470;
Best Local Similarity 92.8%; Pred. No. 2.9e-57;
Matches 324; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 19 ccgctagagagactgcactgcccattatccctgcgacctgcgctccattagggt 78
|||||
DB 386 CCGCTAGAGCGAGACTGCATCGCCATCTCTCACTGCACCTGCGGGTCCCATCAGGGCT 327
|||||
QY 79 gcagctccgctcagatgacgcttagggcgagctcctgcgctgctcctgcggaact 138
|||||
DB 326 CGAGCCTCCGGCTACCATGACGCTGGGGCGCAGCTCTGCGGTGCTGCGGGATCT 267
|||||
QY 139 acacagtagtgggggcggtccacatctgatgagccacagagcgtgtcccccgcca 198
|||||
DB 266 CTACAGCAGTGGCGCGGTGTCAGCTGTGACGCCGCCACGAGACGCTGTCCCGGCCA 207
|||||
QY 199 caggagagccacccctggcctctgtctgtatgtacacgggtccatgtggagcgtctga 258
|||||
DB 206 CAGGAGAGCCACCCCTGCTGCTTCGATGTCAACGGTCCATGTGGGATGACCTGA 147
|||||
QY 259 tgcaggtgatgcagcgccctcagcattctggagcagctctgagcagcgccgcccggg 318
|||||
DB 146 TGCAGGTGATGACGGCGCCCTCGCGCATCTGGAGCGAGCTGTGAGCGCCGACGCCGG 87
|||||
QY 319 tcatcgccaactatgcgctggtgctcttccacgacccagacatggccc 367
|||||
DB 86 TCATCGCCAACTATGCGTGTGCTGCTTTCCAGCAGCCAGCTAGAGCCCC 38
|||||
RESULT 2
AZ985073/c 720 bp DNA GSS 27-APR-2001
LOCUS 2M0266L17R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
DEFINITION clone UUGC2M0266L17 R, DNA sequence.
ACCESSION AZ985073
VERSION AZ985073.1 GI:13856300
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 720)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
```

```
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0266 row: L column: 17
Seq primer: CACACAGGAACAGCATATGACC
Class: plasmid ends
High quality sequence stop: 720.
Location/Qualifiers
1..720
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0266L17"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 203 a 134 c 145 g 238 t
ORIGIN
Query Match 8.8%; Score 90.4; DB 13; Length 720;
Best Local Similarity 72.0%; Pred. No. 8.9e-10;
Matches 118; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 426 caggagagtggtgactgccagaaatgagtgtgggggcccatacaggtgcggtgaggtt 485
|||||
DB 274 CAGGGAGGTGCTGACTGCTCCAGAAATGAGCATCGGGGCTATAAAGATTGCCTTGGAAATC 215
|||||
QY 486 gccaaccccggtcctcatctacgtctcttcggatgccgtgccaaagactaccacaag 545
|||||
DB 214 TCTCTTCTGGGTCTTCATCTATGTTTTCATGATGACCATCCAGGATTTATCGGCTC 155
|||||
QY 546 aagaatgagctcctgcagctcctgcagctgaagcagctgcaggt 589
|||||
DB 154 ACCCATGAGGTGCTGCAGCTTATCCACAGAGAAACAATCTCAGGT 111
|||||
RESULT 3
CNS025DE 937 bp DNA GSS 12-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence PUC-Or1 end of clone
DEFINITION 236N09 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL181931
VERSION AL181931.1 GI:7820017
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 937)
```

RESULT	4
BF418902	
LOCUS	BF418902 559 bp mRNA EST 28-NOV-2000
DEFINITION	UI-R-BJ2-bqk-c-06-0-UI.s1 UI-R-BJ2 Rattus norvegicus cDNA clone
ACCESSION	UI-R-BJ2-bqk-c-06-0-UI 3', mRNA sequence.
VERSION	BF418902
KEYWORDS	BF418902.1 GI:11406891
SOURCE	EST.
ORGANISM	Norway rat. Rattus norvegicus Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sclurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 559) Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery Genome Res. 6 (9), 791-806 (1996)
JOURNAL	9704477
MEDLINE	Contact: Soares, MB
COMMENT	Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA

	Query Match	5.88;	Score 70.6;	DB 11;	Length 559;
	Best Local Similarity	63.08;	Pred. No. 1.8e-05;		
	Matches 109;	Conservative	0;	Mismatches 64;	Indels 0; Gaps 0;
Qy	201	ggagacgccaccctggcctctctcgatgtaccggctccatgtggagcagctctgatg	260		
Db	386	GGGGCGTCCACGTTGGGCTTTTGTCTTTGATGTGACTGGCTCCATGATGATTAGTT	445		
Qy	261	caggtgatcgacggcgctccacgcattctggagcgcagctctgagcgcgcagccgggctc	320		
Db	446	CAGGTTATTGAAGGGGCTTCCAAATTTTGACAGCGTCTTTGAAAGACCTTAAGACCT	505		
Qy	321	atcgccaactatgcgtgtgtgctcttcacagccacagacattggcccaagtac	373		
Db	506	CTTTACAACTTTGCTTTGGTGGCCCTTCCATGATGACAGAAATCGGTTCAGTGAC	558		

RESULT	5				
LOCUS	AZ137837/c				
DEFINITION	<p>SP_0178_A2_E10_SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=178 Col=20 Row=1, DNA sequence.</p>				
ACCESSION	AZ137837	915 bp	DNA	GSS	28-AUG-2000
VERSION	AZ137837.1	GI:8289740			
KEYWORDS	GSS.				
SOURCE	Strongylocentrotus purpuratus.				
ORGANISM	Strongylocentrotus purpuratus				
	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;				
	Echinoidea; Euechinoidea; Echinacea; Echinoida;				
	Strongylocentrotidae; Strongylocentrotus.				
REFERENCE	1 (bases 1 to 915)				
AUTHORS	Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.				
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)				

50 others

```

Query Match          4.88; Score 49.6; DB 13; Length 935;
Best Local Similarity 29.08; Pred. No. 0.64;
Matches 122; Conservative 99; Mismatches 198; Indels 1; Gaps 1;

QY    61 gcgcgtccattaggctcagctccggcttcagcatgaacgacctaggcgagctctctgc 120
      | : : : | : : : | : : | : : | : : | : : | : : | : : | : : |

```

Db 924 GSSGSGCGSGGSGGSCCCCSCGCGCCSCSCSSSSCCSCSGCGCGC 865

Db	864	CSCGGSSCCGCGCGSGSSGCGCCGCGGCGCCGCGGSCCCSSSCCGCGCGSCC	805
Ov	181	aaacgctatccccgcacacgaagacgcacacgctgccttgctcttgatgtcacggct	240

Db 804 GCGCCGCGGCCGGCGSSGGCGSGCKCGCGCGSGCSSSGS 745

QY 241 ccattgggacgatctgatcaggtgatcgacgcgcctcacgcattctggagcgcaatc 300

Db	744	SCGGSGGSGCS	SSGCGC	-GSSGCGSGCGCCGCGCGCGSGCGCGCGSGS	686
Qy	301	tgagcagcgcgagccgggtatcgcgaactatgcgctgggtgccttccagacccagaca			360

DB 685 CGCGCGCCGCCCCCSCSSSFGSSCSGCCSSSCSSSCSSSGCCGGCCCSCGCG 626

QY 361 ttggccagtgacctcaacgcggaaccagtggtttcagagagagtgaacaactct 420

Qy 421 atgttcaggagtgtagtgcacgaaaatgagtgtggggcccatcaaggctgccgtgg 480
| : : | : | : | : : : : | : : | : | : | :
Db 565 GMGCRAGSKWGSAGSSRCGCACSGSGSBSKRRKGASGCASCSSAYKSGSCGGB 506

LOCUS	925 bp	DNA	GSS	03-JUN-1999
CNS0091P				
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC #			

ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.

ORGANISM
Drosophila melanogaster
Eukarya; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
TITLE
JOURNAL
1 (bases 1 to 925)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial

isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	SOURCE
1. High Accuracy: The model achieves a high accuracy rate, consistently performing well across various datasets and tasks.	Model Performance Metrics
2. Scalability: The system is designed to handle large volumes of data and complex tasks, ensuring performance remains stable as the workload increases.	System Architecture
3. Interpretability: The model's decisions are transparent and explainable, allowing users to understand the underlying factors influencing the results.	Model Explainability
4. Robustness: The system is resilient to noise and outliers, maintaining high performance even in less-than-ideal conditions.	Model Robustness
5. Efficiency: The model is optimized for fast processing times, ensuring timely results for time-sensitive applications.	Model Efficiency
6. Flexibility: The system can be adapted to different environments and tasks, providing a versatile solution for various use cases.	Model Flexibility
7. Integration: The model seamlessly integrates with existing systems and workflows, facilitating easy adoption and implementation.	System Integration
8. Security: The system implements robust security measures to protect data and ensure compliance with relevant regulations.	System Security
9. Customization: The model can be tailored to specific user requirements, providing a personalized experience.	Model Customization
10. Support: Comprehensive documentation and user support are provided to assist in the deployment and maintenance of the system.	User Support

```

1. .925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR19D16"
/note="end : TET3"

```

BASE COUNT	120 a	61 c	61 g	172 t	511 others
ORIGIN					

Query Match 4.7%; Score 48.8; DB 13; Length 925;
Best Local Similarity 15.5%; Pred. No. 0.96;
Matches 60; Conservative 167; Mismatches 158; Indels 1; Gaps 1;

[illegible]

RESULT 10
RG895162

LOCUS	BC895162	520 bp	mrna	EST	05-JUN-2001
DEFINITION	38378 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.				
ACCESSION	BC895162				
VERSION	BC895162.1	GI:14305403			
KEYWORDS	EST.				
ORGANISM	plg.				
SCROFA	Sus scrofa				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
AUTHORS
1 (bases 1 to 520)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
TOURNAT Unpublished 2000

UNPUBLISHED (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904 e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 121 row: M column: 3
Seq primer: ATTTAGGTGACACTATAG.

FEATURES

```
1. 520
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="PH10B"
```

```

/note=Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
95 a 163 a 98 r
BASE COUNT

```

BASE COUNT	95 a	164 c	163 g	98 t
ORIGIN				

Query Match	4.5%;	Score 46.6;	DB 11;	Length 520;
Best Local Similarity	51.2%;	pred. No. 2.9;		
Matches 109;	Conservative 0;	Mismatches 104;	Indels 0;	Gaps 0;
QY	108	gcgcagctctcgcgctgctctctggcgacctacacagtagtgcgcgcgcggtcacatct	167	
Db	150	GCCCACCGGTACGCGTGTCTCTCGACACCTGCGAGCAGCTGTGAGGCTGCCCTGCTTGGAT	209	
QY	168	gatgagcccaacgaagcagctgtctcccccgcacagggagagcaccacctggctcttc	227	
Db	210	GTGCGACTTCAACAAAGGCCAAGTCCCCGCTACCAAGAGATGCCACACAGCACATAGGCTGC	269	
QY	228	gatgtcacgcggtccatgtgtgggcagctatgatcagctgatcagcgcgctcagcatt	287	
Db	270	GTCTTCTGCGACCCCATCTTCCAGGAGCTGCTGGAGCTGAAGCATGCGAGCCTCTCTGGAC	329	
QY	288	ctggagcgcagctgtgagccgcagcccgggtc	320	
Db	330	CTATACCGTGGGCCCGGAGGAGGAGGCG	362	

RESULT 11
CNS0091P/c

LOCUS	CNS0091P	925 bp	DNA	GSS	03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR1916 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL053013				
VERSION	AL053013.1	GI:4934461			
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 925)				

REFERENCE AUTHORS

TITLE
Direct Submission

SUBMITTED (02-JUN-1999)

JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT
Determination of this BAC-end and sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial,
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or

008El6	G from Tetraodon nigroviridis, genomic survey sequence.
AL234371	GI:7893506
AL234371.1	GI:7893506
GSS:	genome survey sequence.
Tetraodon nigroviridis	
Tetraodon nigroviridis	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.	
1 (bases 1 to 818)	
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier.F., Saurin,W., Bernot,A. and Weissenbach,J.	
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	
Unpublished	
2 (bases 1 to 818)	
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.	
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence	
Unpublished	
3 (bases 1 to 818)	
Genoscope.	
Direct Submission	
Submitted (12-APR-2000)	to the EMBL/GenBank/DBDJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .	
Location/Qualifiers	
1..818	

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone="008E16"

/clone_lib="G"

BASE COUNT 142 a 104 c 245 g 149 t 178 others

Query Match 4.4%; Score 45; DB 13; Length 818;
Best Local Similarity 25.9%; Pred. No. 6.4;
Matches 87; Conservative 101; Mismatches 143; Indels 5; Gaps 1;
QY 187 tttccccccacagagagcgcacccctgcttgcgttcttctgctgacccgctccatgt 246
Db 622 TCTCTCTMAACACTCATYTTTCACCTTTCCCTCTCTCATCTCCCMCCCTTCTCYTTC 563
QY 247 gggagatctgagcaggtgagcggcgctcagcagcttctgagcagcagctgagca 306
Db 562 CCCACAAACATAAASST 503
QY 307 ggcgagcgggtcatcgcgaactatgcgtggtgcttccacagcagcagcattgccc 366
Db 502 TSTSSSBTTATSSST 443
QY 367 cagtgacctcagcgggacccagtggtgttccagagaga-----gctgagacaaactcta 421
Db 442 ASATSSASASASS 383
QY 422 tttcagagagctgactgcagcagaaatgagtgtggggccatcaaggctgcgcgtgga 481
Db 382 ASSNCAGSSASS 323
QY 482 ggttgcaacccggctctctcatctacgtcttctc 517
Db 322 GSSGCGAGAGSSCTCTACASSSCTSSSSASSTSTC 287

RESULT 14
LOCUS BF481079 375 bp mRNA EST 05-DEC-2000
DEFINITION FM1_16_B06.bl_A003 Floral-Induced Meristem 1 (FMI) Sorghum
propinquum cdna, mRNA sequence.
ACCESSION BF481079
VERSION BF481079.1 GI:11551900
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 375)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
L.H.

TITLE An EST database from Sorghum: floral-induced meristems
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 342
POLYA-No. Location/Qualifiers
1..375
/organism="Sorghum propinquum"
/db_xref="taxon:132711"

FEATURES
source
1..375
/organism="Sorghum propinquum"
/db_xref="taxon:132711"

/clone_lib="Floral-Induced Meristem 1 (FMI)"
/note="Organ: Floral-Induced Meristems; Vector:
pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; mature plants were placed in a growth chamber for
15 days with 16 hr darkness and 8 hr light (flowering is
induced by short-day conditions); 16 days after being
returned to the greenhouse under natural long days during
late April/early May, meristems were harvested The
library was made from poly-A RNA in the cloning vector
lambda Zap II. Clones to be sequenced were prepared by
mass excision."

BASE COUNT 75 a 103 c 133 g 63 t 1 others
ORIGIN
Query Match 4.3%; Score 44.8; DB 11; Length 375;
Best Local Similarity 52.7%; Pred. No. 7.3;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 560 gcagctctctgcagctgaagcagctcaggtgtcttctgctgactggggaactcggtga 619
Db 91 GCAGCAGCGCACCTGGCTGGCGCACCAAGCTGTCTTCTGTCTGCGCTCCGACGTCCA 150
QY 620 ccgcacccaccctgctacctgcttggagagatcgctccacagcttctggccaagt 679
Db 151 CGACCTCGCAAGGTCGACCTACGCGCTGACGTGCTCGGGCAGTCAATCCGATGACAT 210
QY 680 gttcagctggaagcagcagcaggtgtcgagggtgttaagtggtgagtcgcgcaccca 739
Db 211 GCGGCTCTGTACGAGTGTGCTGGCGGCGGACGGCTACTGGAGATGGACGCGCGCTGT 270
QY 740 gccc 743
Db 271 GGCC 274

RESULT 15
LOCUS BG560654 379 bp mRNA EST 10-APR-2001
DEFINITION RH122_58_B04.bl_A003 Rhizome2 (RH122) Sorghum propinquum CDNA, mRNA
sequence.
ACCESSION BG560654
VERSION BG560654.1 GI:13589652
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 379)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
L.H.

TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 330
POLYA-No. Location/Qualifiers
1..379
/organism="Sorghum propinquum"
/db_xref="taxon:132711"

FEATURES
source
1..375
/organism="Sorghum propinquum"
/clone_lib="Rhizome2 (RH122)"
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda
zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

Search completed: March 7, 2002, 21:07:19
Job time: 8057 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 21:07:19 ; Search time 1634.39 seconds
(without alignments)
5424.206 Million cell updates/sec

Title: US-09-665-728-2_COPY_96_920
Perfect score: 825
Sequence: 1 atgagcctagggcagct.....ctcctcttctgatgcaagac 825

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_esti:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	240	29.1	470	10	AI449932 mr81b03.x
C 2	90.4	11.0	720	13	AZ985073
C 3	77.4	9.4	937	13	CNS025DE
C 4	70.6	8.6	559	11	BF418902 UI-R-BJ2-
C 5	64	7.8	915	13	AZ137837
C 6	61	7.4	528	10	BE508854 del3h07.y
C 7	55	6.7	754	13	AZ189426 SP.1014.A
C 8	46.6	5.6	520	11	BC895162 358378 MA
C 9	46.2	5.6	935	13	AL066051 Drosophil
C 10	46	5.6	965	13	AL178322 Tetraodon
C 11	45	5.5	818	13	CNS039U2
C 12	44.8	5.4	375	11	BF481079 FMI_16_B0

13	44.8	5.4	379	11	BG560654
14	44.8	5.4	392	11	BF481089
15	44.8	5.4	406	11	BG560698
16	44.2	5.4	925	13	CNS0091P
17	43.2	5.2	523	11	BG240204
18	43.2	5.2	924	13	CNS04SRD
19	42.2	5.1	461	10	AU174537
20	42	5.1	632	11	BG744240
21	41.8	5.1	503	10	AI653725
22	41.8	5.1	686	11	BG835990
23	41.8	5.1	925	13	CNS0091P
24	41.4	5.0	384	10	AA294693
25	41.4	5.0	476	10	BE477180
26	41.4	5.0	566	10	AW082912
27	41.4	5.0	872	11	BG345072
28	41	5.0	647	11	BG180123
29	41	5.0	778	10	AI991270
30	41	5.0	814	10	AW081400
31	40.8	4.9	523	10	AI855212
32	40.8	4.9	556	11	BG874151
33	40.8	4.9	564	13	AQ848793
34	40.8	4.9	586	10	BE553066
35	40.8	4.9	603	10	AI855218
36	40.8	4.9	626	10	AI855245
37	40.8	4.9	671	13	AQ901789
38	40.8	4.9	813	11	BG321263
39	40.8	4.9	956	13	CNS03GMI
40	40.6	4.9	460	10	AI432376
41	40.6	4.9	738	11	BG864336
42	40.4	4.9	459	10	AW504911
43	40.4	4.9	470	11	BF133055
44	40.4	4.9	472	10	AI330202
45	40.4	4.9	503	11	BG754939

ALIGNMENTS

RESULT 1

AI449932/c

LOCUS

DEFINITION

mr81b03.x1 Stratagene mouse heart (#937316)

IMAGE:603821 3' similar to WP:F15G9.4 CE01552 IG SUPERFAMILY

REPEATS:,, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AI449932 470 bp mRNA EST 09-MAR-1999
mr81b03.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
IMAGE:603821 3' similar to WP:F15G9.4 CE01552 IG SUPERFAMILY
REPEATS:,, mRNA sequence.
AI449932.1 GI:4293462
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 470)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LINL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone: polyt not found
High quality sequence stop: 467.
Location/Qualifiers
1. 470
source

/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone_image=603821"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="j13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: heart; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. 93 pooled NIH/Swiss j13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR vector; -5'
adaptor sequence: 5' GAATTCGGCAGGAC 3' adaptor sequence: 5' CTCGAGTTTTTTT TTTT 3'
BASE COUNT 80 A 152 C 168 G 69 T 1 others
ORIGIN
Query Match 29.1%; Score 240; DB 10; Length 470;
Best Local Similarity 92.6%; Pred. No. 1.le-44;
Matches 252; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 1 ataacgctaggcgacgtcctgcggctccctgccgtacctcacagttagtgcgcgcg 60
Db 309 ATCACGCCTCGGGCGCAGCTCTGCCGTCTGCTGGCGCATCTACAGCAGTGCGCGCG 250
QY 61 gcggtcacattgatgaccagcaagaacactgtcccccccacaggaagacccaacctg 120
Db 249 GTGGTCACTGTTCAGCCGCCCCAAGAACAACACTGTGCCCACCACAGGAGACGCCATA 190
QY 121 gcccttccttcgatgtcaacggctcatctggaacatctgtgcagggtgatacgacg 180
Db 189 GCCTTCGTCATGATGTCACCGGCTCTCATGTGGATGACCTGATGACAGGTGATCGAGCG 130
QY 181 gcttacgcattttggagcagcttgtgacagcgcgcgggtcatcgaccaactatgcy 240
Db 129 GCCTCGCGCATCTTGGAGCGCAGCTTAGCAGCAGCGCGCATTGCGCAACTATGCG 70
QY 241 ctagtgcctttccagaccacagacattggccc 272
Db 69 CTGCTGCTTTCCACGACCCAGGTAGAGCCCC 38
RESULT 2
LOCUS AZ985073/c
DEFINITION ZM0266LL17R Mouse 10kb plasmid UUGCM library Mus musculus genomic clone UUGC2M0266L17 R, DNA sequence.
ACCESSION AZ985073
VERSION AZ985073.1 GI:13856300
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 720)
Dunn,D., Ayvagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly, .M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D..Weiss,R.
Title Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Journal Unpublished (2000) Comment Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunne@genetics.utah.edu Insert Length: 10000 Std Error: 0.00

is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 227-278, >(GAAAA)n\$Simple_repeat (Seq primer: M13 Forward POLYA=No. Location/Qualifiers 1..559 /organism="Rattus norvegicus" /strain="Sprague-Dawley" /db_xref="taxon:10116" /clone="UI-R-BJ2-bqk-c-06-0-UI" /clone_lib="UI-R-BJ2" /lab_host="DH10B (Life Technologies)" /note="vector: pT7T3D-pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2

143 a 137 c 152 g 127 t

[illegible]

AZ137837 915 bp DNA GSS 28-AUG-2000
SP_0178_A2_E10_SP6E Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library Strongylocentrotus purpuratus
genomic clone Plate=178 Col=20 Row=I, DNA sequence.
AZ137837
AZ137837.1 GI:8289740
GSS.
Strongylocentrotus purpuratus.
Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinozoa; Echinacea; Echinoidea;
Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 915)
Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,
Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray
G.A., Etensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and
Hood, L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology

ORIGIN

Query Match 5.6%; Score 46.6; DB 11; Length 520;
 Best Local Similarity 51.2%; Pred. No. 1.1;
 Matches 109; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 13 qcagcgtcctgcgcgtcctctggcgacacagtagtggcgcgggcggtcacatct 72
 Db 150 GCCACCGTGTACCGGTGTCTCCGACACCTGCAGCACCTGTGAGGCTGCCTGCTGAT 209

QY - 73 gatgagcccaagacgcgtgtccccccacagagacgccacccttgcgtcttc 132
 Db 210 GTCGACTTCAACAGGCCAAGTCCCGCTACACGAGATGCCACAGCATAGGCTGC 269

QY 133 gatgtaccgagctccatgtggagcatctgatgcaggtgatcgacgcgcctcacgatt 192
 Db 270 GTCTTCTGGACCCATCTTCCAGAGCTGTGTGAGCTGAAGCATGCAGCCTCTCGCCAC 329

QY 193 ctggagcgcagctgtgagcgcgcgcgcgcgggtc 225
 Db 330 CCTATGCACCGTGGCGCCGGAGGAGGCAGGCC 362

RESULT 9
 CNS006XK/c
 LOCUS
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL066051
 VERSION AL066051.1 GI:4945019
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 935)
 AUTHORS Direct Submission
 TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
 COMMENT - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1..935
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR14N09"
 /note="end : T7"

BASE COUNT 257 a 170 c 162 g 96 t 250 others
 ORIGIN

Query Match 5.6%; Score 46.2; DB 13; Length 935;
 Best Local Similarity 29.4%; Pred. No. 1.5;
 Matches 112; Conservative 90; Mismatches 178; Indels 1; Gaps 1;


```

QY 465 gcagctctgcagctgaagcagctgcaggggtgtcttctgctgactgaggggactgcggtga 524
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 114 GCAGCAGCCGACCTGTGCTGGCGCACAAAGCTGTTCTTCTGTCTGCGCTCCGACGTCGA 173

QY 525 ccgcaccacccctgctacactggcttttggagagatcgctccaccagttcttgccaaagt 584
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 174 CGACCTCGCCAAAGTCGACCTACCGCTGACGTGCTGCGCGCAGTCAAAATCCGATGACAT 233

QY 585 gttccagctggacaagcagcaggtgtgcggaggtgttaaagtgggtggagtcgcgccatcca 644
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 234 GCGCGCTCTGTACGAGTCGCTGGCGGCGGACGCGTACTGGAGATGGACGCGCGCTGCT 293

QY 645 gccc 648
    |||||
Db 294 GGCC 297

RESULT 14
BF481089
LOCUS BF481089 392 bp mRNA EST 05-DEC-2000
DEFINITION FM1_16.C06.bl.A003 Floral-Induced Meristem 1 (FM1) Sorghum
ACCESSION BF481089
VERSION BF481089.1 GI:11551910
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 392)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
,L.H.
TITLE An EST database from Sorghum: floral-induced meristems
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 350
POLYA=No.
FEATURES
    source
    Location/Qualifiers
        1..392
            /organism="Sorghum propinquum"
            /db_xref="taxon:132711"
            /clone_lib="Floral-Induced Meristem 1 (FM1)"
            /note="Organ: Floral-Induced meristems; Vector:
            pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
            EcoRI; mature plants were placed in a growth chamber for
            15 days with 16 hr darkness and 8 hr light (flowering is
            induced by short-day conditions); 16 days after being
            returned to the greenhouse under natural long days during
            late April/early May, meristems were harvested The
            library was made from poly-A RNA in the cloning vector
            lambda ZAP II. Clones to be sequenced were prepared by
            mass excision."
BASE COUNT 79 a 108 c 135 g 68 t 2 others
ORIGIN

Query Match 5.4%; Score 44.8; DB 11; Length 392;
Best Local Similarity 52.7%; Pred. No. 2.6;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 465 gcagctctgcagctgaagcagctgcaggggtgtcttctgctgactgaggggactgcggtga 524
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 91 GCAGCAGCCGACCTGTGCTGGCGCACAAAGCTGTTCTTCTGTCTGCGCTCCGACGTCGA 150

QY 525 ccgcaccacccctgctacactggcttttggagagatcgctccaccagttcttgccaaagt 584
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 151 CGACCTCGCCAAAGTCGACCTACCGCTGACGTGCTGCGCGCAGTCAAAATCCGATGACAT 210

QY 585 gttccagctggacaagcagcaggtgtgcggaggtgttaaagtgggtggagtcgcgccatcca 644
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 211 GCGCGCTCTGTACGAGTCGCTGGCGGCGGACGCGTACTGGAGATGGACGCGCGCTGCT 270

QY 645 gccc 648
    |||||
Db 271 GGCC 274

RESULT 15
BG560698
LOCUS RHIZ2_58_F04.bl.A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
DEFINITION RHIZ2_58_F04.bl.A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
ACCESSION BG560698
VERSION BG560698.1 GI:13589696
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 406)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 386
POLYA=No.
FEATURES
    source
    Location/Qualifiers
        1..406
            /organism="Sorghum propinquum"
            /db_xref="taxon:132711"
            /clone_lib="Rhizome2 (RHIZ2)"
            /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
            Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
            from poly-A RNA in the cloning vector lambda ZAP II.
            Clones to be sequenced were prepared by mass excision."
BASE COUNT 90 a 101 c 151 g 64 t
ORIGIN

Query Match 5.4%; Score 44.8; DB 11; Length 406;
Best Local Similarity 52.7%; Pred. No. 2.6;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 465 gcagctctgcagctgaagcagctgcaggggtgtcttctgctgactgaggggactgcggtga 524
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 59 GCAGCAGCCGACCTGTGCTGGCGCACAAAGCTGTTCTTCTGTCTGCGCTCCGACGTCGA 118

QY 525 ccgcaccacccctgctacactggcttttggagagatcgctccaccagttcttgccaaagt 584
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119 CGACCTCGCCAAAGTCGACCTACCGCTGACGTGCTGCGCGCAGTCAAAATCCGATGACAT 178

QY 585 gttccagctggacaagcagcaggtgtgcggaggtgttaaagtgggtggagtcgcgccatcca 644
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 179 GCGCGCTCTGTACGAGTCGCTGGCGGCGGACGCGTACTGGAGATGGACGCGCGCTGCT 238

QY 645 gccc 648

```

Db 239 GCC 242
||||

Search completed: March 7, 2002, 21:07:21
Job time: 8059 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: March 7, 2002, 20:48:02 ; Search time 48.12 Seconds
(without alignments)
423.320 Million cell updates/sec

Title: US-09-665-728-1
Perfect score: 1403
Sequence: 1 MTPRAQLPLLATYTWAA.....IEVRDLGMSQSGPPLLMQD 275

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_ll01.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1403	100.0	275	22 AAB83147	Rat secreted facto
2	628.5	44.8	3117	21 AAV53667	Sequence g1/32818
3	116	8.3	353	22 AAB98973	Murine SHD1 protei
4	98	7.0	1017	22 AAB86124	Tobacco formylglyc
5	93	6.6	750	22 AAB48252	Soybean magnesium
6	92.5	6.6	414	13 AAR20202	S.equisimilis stre
7	91.5	6.5	369	12 AAR12892	Truncated Met stre
8	91.5	6.5	372	21 AAY84006	Amino acid sequenc
9	91.5	6.5	384	21 AAY84007	Amino acid sequenc
10	91.5	6.5	401	21 AAY84004	Amino acid sequenc
11	91.5	6.5	413	21 AAY84005	Amino acid sequenc

12	91.5	6.5	414	20 AAB86143	Streptokinase (SK)
13	91.5	6.5	414	20 AAB86144	De-immunised strep
14	91.5	6.5	414	21 AAB01295	Wild type streptokin
15	91.5	6.5	414	21 AAB01296	Altered streptokin
16	91.5	6.5	440	12 AAR12889	Streptokinase. St
17	91.5	6.5	483	12 AAR12885	Factor Xa-cleavabl
18	91.5	6.5	483	12 AAR12522	Factor Xa-cleavabl
19	91.5	6.5	499	12 AAR12891	Streptokinase fuse
20	91.5	6.5	747	12 AAR12894	Met-core streptoki
21	91.5	6.5	859	12 AAR12893	OmpAL streptokinas
22	91.5	6.5	1102	19 AAR23948	Porcine phosphoino
23	91.5	6.5	1102	20 AAW90083	Porcine G-protein
24	91.5	6.5	1102	21 AAY76802	Pig p120 regulator
25	90.5	6.5	693	20 AAY30796	A mouse transgluta
26	89.5	6.4	755	22 AAB48248	Corn magnesium che
27	89	6.3	800	18 AAW21723	Plasminogen-bindin
28	89	6.3	813	18 AAW21728	Wild type plasmino
29	89	6.3	1181	18 AAW21727	Streptokinase/malt
30	89	6.3	1194	18 AAW21724	Modified streptoki
31	89	6.3	1194	18 AAW21726	Streptokinase/malt
32	89	6.3	1194	18 AAW21725	Modified streptoki
33	88.5	6.3	404	22 AAG91953	C glutamicum prote
34	88.5	6.3	440	11 AAR06377	Streptokinase G pr
35	88.5	6.3	511	16 AAR69506	Aspergillus sp. re
36	88.5	6.3	515	16 AAR69505	Aspergillus sp. re
37	88.5	6.3	515	16 AAR69504	Aspergillus sp. re
38	88.5	6.3	515	22 AAB49946	A. oryzae protein
39	88.5	6.3	1407	22 AAB86123	A. thaliana formyl
40	88	6.3	848	21 AAB43716	Human cancer assoc
41	88	6.3	896	17 AAW00733	Human major vault
42	87	6.2	758	19 AAW81771	Tobacco CHLD prote
43	86.5	6.2	348	12 AAR10196	Streptokinase (1-3
44	86.5	6.2	356	20 AAY24795	Recombinant strept
45	86.5	6.2	372	12 AAR10197	Streptokinase (1-3

ALIGNMENTS

RESULT 1					
AAB83147					
ID	AAB83147	standard; protein; 275 AA.			
XX	AC	AAB83147;			
XX	DT	29-JUN-2001 (first entry)			
XX	DE	Rat secreted factor encoded by clone P00210D09.			
XX	KW	Rat; secreted factor; P00210D09; cardiant; nephrotropic;			
KW	KW	antiinflammatory; gene therapy; cardiac disease; renal disease;			
KW	KW	inflammatory disease.			
OS	Rattus sp.				
XX	Key	Location/Qualifiers			
FT	Peptide	1..21			
FT	Protein	/label= Signal_peptide			
FT	Domain	22..275			
FT	Domain	/note= "Rat secreted factor"			
FT	Domain	35..55			
FT	Domain	/label= Transmembrane_domain			
FT	Domain	123..143			
FT	Domain	/label= Transmembrane_domain			
XX	WO200123419-A2.				
XX	05-APR-2001.				
XX	27-SEP-2000; 2000WO-US26582.				
XX	27-SEP-1999; 99US-0156277.				

PA (SCIO-) SCIOS INC.
 XX Stanton LW, Kapoun AM;
 XX WPI; 2001-328177/34.
 XX N-PSDB; AAF82464.
 XX Novel secreted factor encoded by clone P00210D09 useful for diagnosing,
 PT treating and/or preventing various cardiac, renal and inflammatory
 PT diseases -
 XX
 XX Claim 9; Fig 1; 69pp; English.
 XX
 CC The present sequence is a novel secreted factor encoded by rat cDNA
 CC clone P00210D09. The invention relates to a polypeptide comprising a
 CC sequence of at least 80% identity to residues 22-122 of the present
 CC sequence, or a sequence encoded by a nucleic acid hybridizing under
 CC stringent conditions to the complement of the coding region comprising
 CC 1031 nucleotides, and having at least one biological activity of the
 CC polypeptide encoded by clone P00210D09. The polypeptides and
 CC polynucleotides of the invention are useful for the treatment of
 CC cardiac, renal and inflammatory diseases. The polynucleotides are
 CC useful in antisense mediated gene inhibition and in gene therapy.
 CC The polypeptides are useful in assays for identifying lead compounds
 CC that may be used as therapeutic agents in the treatment of cardiac,
 CC kidney or inflammatory diseases.
 XX
 SQ Sequence 275 AA;
 Query Match 100.0%; Score 1403; DB 22; Length 275;
 Best Local Similarity 100.0%; Pred. No. 2.2e-138;
 Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTPRAQLLLATYVAAAVTSDEPTKTLSPATGDTAFVFDVTGSMWDDLMQVIDG 60
 DB 1 MTPRAQLLLATYVAAAVTSDEPTKTLSPATGDTAFVFDVTGSMWDDLMQVIDG 60
 QY 61 ASRILERSLSRSRVIANVALYVPHDPDIPVTLTADPVVFORLQLYVGGDCPEMS 120
 DB 61 asrilerslsrsrvianvalyvpfhdpgvlttdpvpvfqrelrqlvvgggdcpems 120
 QY 121 VCAIKAAVEANPGSFIYVFSARAKDYHKKNELQLQLKQSOVVFVLTGCGDRTHPG 180
 DB 121 vgaikaaveanpgsfliyfsdarakdyhknellqlqlkqsgvvfvltgdcgdrthpg 180
 QY 181 YIAFETIASTSSGOVFQDKQOVSEVLKQVSAIOASKVHLLSADHEEGEHTWRIPFDP 240
 DB 181 ylafeiaastsggvfqlkqkvsevlkwvesaioaskvhllsadheeegehtwrifpdp 240
 QY 241 SLKEVTISLGGPPEIEVRDPLGMSQGSPPLLMQD 275
 DB 241 slkevtislspppeievrplgmsgsgsppllmqd 275
 RESULT 2
 AAY33667
 ID AAY33667 standard; Protein; 3117 AA.
 XX
 AC AAY33667;
 XX
 DT 22-FEB-2000 (first entry)
 XX
 DE Sequence gi/3328186 from an alignment with protein 608.
 XX
 KW Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
 KW bone development; gi/3328186.
 XX
 OS Unidentified.
 XX
 PN WO9960164-A1.
 XX
 DT 25-NOV-1999.
 PD

XX 14-MAY-1999; 99WO-US11066.
 XX 15-MAY-1998; 98US-0085673.
 XX (QUAR-) QUARK BIOTECH INC.
 XX Einat P, Mor O, Skalliter R, Feinstein E, Faerman A;
 XX WPI; 2000-053304/04.
 XX
 PT Identification of stress induced genes for determining risk and
 PT preventing, treating or controlling osteoporosis -
 XX
 XX Claim 32; Fig 6A-R; 308pp; English.
 XX
 CC The present sequence is obtained from a clustral X alignment with
 CC protein 608. Protein 608 was identified using the method of the invention
 CC after subjecting rat osteoblasts to mechanical stress. Expression of the
 CC 608 gene was found to be upregulated by about 3-fold in cells subjected
 CC to mechanical strain. The specification describes a method for the
 CC identification of genes responsive to a specific mechanical stress. The
 CC method comprises applying the mechanical stress to an organism (tissue
 CC or cells comprising bone cells), isolating the specific cellular
 CC fractions and extracting mRNA from them, and differentially analysing the
 CC mRNA in comparison with control samples. The method is used to identify
 CC genes whose expression is responsive to a specific stress. The identified
 CC genes are employed in determining risk associated with a physiological or
 CC disease state. The risk determination methods are used for testing a
 CC medicament for gene therapy. These medicaments, or genes identified by
 CC the method of the invention, are used for treating, preventing or
 CC controlling a physiological or disease state (especially osteoporosis or
 CC bone density or other factors causing or contributing to osteoporosis or
 CC its symptoms or other conditions involved in mechanical stress or its
 CC lack. The methods can also be used for advancing research or studies in
 CC bone development.
 XX
 SQ Sequence 3117 AA;
 Query Match 44.8%; Score 628.5; DB 21; Length 3117;
 Best Local Similarity 48.6%; Pred. No. 9.5e-56;
 Matches 126; Conservative 51; Mismatches 69; Indels 13; Gaps 3;
 QY 10 LLLATYVVAAVTSDEPTKTLSPATGDTAFVFDVTGSMWDDLMQVTDGASRILERSL 69
 DB 10 LLLATYVVAAVTSDEPTKTLSPATGDTAFVFDVTGSMWDDLMQVTDGASRILERSL 69
 QY 15 LLLATYVVAAVTSDEPTKTLSPATGDTAFVFDVTGSMWDDLMQVTDGASRILERSL 66
 DB 15 LLLATYVVAAVTSDEPTKTLSPATGDTAFVFDVTGSMWDDLMQVTDGASRILERSL 66
 QY 70 SRSRVIANYALVPHDPDIPVTLTADPVVFORLQLYVGGDCPEMSVGAATKAAVE 129
 DB 70 SRSRVIANYALVPHDPDIPVTLTADPVVFORLQLYVGGDCPEMSVGAATKAAVE 129
 QY 67 adrekliynyimvpfhdpgvlttdpvpvfqrelrqlvvgggdcpektitgllkaiq 126
 DB 67 adrekliynyimvpfhdpgvlttdpvpvfqrelrqlvvgggdcpektitgllkaiq 126
 QY 130 VANPGSFIYVFSARAKDYHKKNELQLQLKQSOVVFVLTGCGDRTHPGYLAPEEIAS 189
 DB 130 vanpgsfliyfsdarakdyhknellqlqlkqsgvvfvltgdcgdrthpgylyapeeias 189
 QY 127 islpssfiyftdarskdyhledvintikeqsgvvfvtgdcgdrthpgftrtyekiaa 186
 DB 127 islpssfiyftdarskdyhledvintikeqsgvvfvtgdcgdrthpgftrtyekiaa 186
 QY 190 TSSGOVFQDKQOVSEVLKQVSAIOASKVHLLSADHEEGEHTWRIPDPSLKEVTISL 249
 DB 190 tssgovfqlkqkvsevlkwvesaioaskvhllsadheeegehtwrifpdpsslkevtsil 249
 QY 187 asfggvfhlekadvstleyvrhvkqkvhlmyearerggtvsvrtnipdvkhlseitia 246
 DB 187 asfggvfhlekadvstleyvrhvkqkvhlmyearerggtvsvrtnipdvkhlseitia 246
 QY 250 SG-----PGPEIEVRDPLG 263
 DB 247 sgkdksdndldiviropeg 265
 RESULT 3
 AAB98973
 ID AAB98973 standard; protein; 353 AA.
 XX
 AC AAB98973;
 XX
 DT 20-AUG-2001 (first entry)
 XX

XX OS Glycine max.
XX PN WO200075340-A2.
XX PD 14-DEC-2000.
XX PF 02-JUN-2000; 2000WO-US15351.
XX PR 04-JUN-1999; 99US-0137461.
XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX PI Butler KH, Famodu OO, Gutteridge S, Maxwell CA;
XX DR N-PSDB; AAC84571.
XX PT Isolated nucleic acid fragments encoding magnesium chelatase subunits,
XX PT useful as probes for genetic and physical mapping of genes, as markers
XX PT for traits linked to these genes, and in plant breeding -
XX PS Claim 10; Page 60-62; 103pp; English.
XX CC The invention relates to nucleic acid fragments encoding magnesium
XX CC chelatase subunits. The nucleic acid fragments may be used to create
XX CC transgenic plants in which the new polypeptides are present at higher or
XX CC lower levels than normal or in cell types or developmental stages in
XX CC which they are not normally found, and for overexpression in bacterial
XX CC or yeast hosts to efficiently produce large amounts of the encoded
XX CC polypeptides which could then be used for screening different compounds
XX CC for potential herbicidal activity. The polynucleotides may also be used
XX CC as probes for genetic and physical mapping the genes that they are part
XX CC of, and as markers for traits linked to these genes. Such information
XX CC is useful in plant breeding. The polypeptides are used for preparing
XX CC antibodies, which are useful for detecting the polypeptides in situ or
XX CC in vitro, and as a target to facilitate design and/or identification of
XX CC inhibitors of enzymes that may be used as herbicides. Host cells may also
XX CC be used directly for screening different compounds for potential
XX CC herbicidal activity. The present sequence represents the soybean
XX CC magnesium chelatase subunit.
XX SQ Sequence 750 AA;

Query Match 6.6%; Score 93; DB 22; Length 750;
Best Local Similarity 23.0%; Pred. No. 1.2;
Matches 54; Conservative 38; Mismatches 85; Indels 58; Gaps 11;

QY 24 SDEPTKTLSPATGDTAFVFDVTGSMWDDLMQVIDGAS-RILERSLSRSRVIANALV 82
DB 533 tdmrakrmakag-alvifvdasgsmalnrmqnakgaalkllaesytrdqv---sii 587
QY 83 PFHDPDGPVTL--TADPVVFORLRLQLYVGGGDCPEMSVGAIAAEEVANPVS---F 136
DB 588 pfirg-daaevllppsrslamarkrierlpcggsgplahgtltavrvinaeksgdgvrm 646
QY 137 IYVFSARAK-----DYHK-----KNELLQL-----LQLKQSQV 165
DB 647 lvaiddgranislkrstdpevaatdapksaqelkdeillevagkiyagmsllvidten 706
QY 166 VFVLGDCGDRTHPGYLAPEEIASTSSGQVFDLDKQ---VSEVLKWEESAIOAS 217
DB 707 kfvtstgfa-----kelarvaqgkyylpnasdavissatkealsalkss 750

RESULT 6
AAR20202
ID AAR20202 standard; Protein; 414 AA.
AC AAR20202;
XX AAR20202;
DT 21-APR-1992 (first entry)

XX DE S.equisimilis streptokinase.
XX KW plasminogen activator; coronary thrombosis; ATCC-9542.
XX OS Streptococcus equisimilis.
XX PN AU9178101-A.
XX PD 28-NOV-1991.
XX XX 31-MAY-1991; 91AU-0078101.
XX PF 23-MAY-1990; 90CU-00000090.
XX PR (INGI-) CENT ING GENETICA B.
XX PA Garcia MPE, Chaplen RR, Hidalgo AP, Doce RS, Marrero LFH;
XX PI Collazo PR, Ramirez AC, Munoz EAM, Martinez WB, Somavilla MC;
XX PI Fernandez AP, Garcia J, Martinez LSH;
XX XX WPI: 1992-024716/04.
XX DR N-PSDB; AAQ20665.
XX PT Streptokinase C-2 gene from S.equisimilis type C - plasmids and
XX PT transformants for large scale intra- and extracellular expression
XX PT of SKC-2 useful in thrombolytic agents
XX PS Claim 13; Page 14; 28pp; English.
XX CC The SKC-2 gene was isolated from S.equisimilis type C by gene
XX CC amplification from the synthetic primers SK1, SK2 and SK3 (see
XX CC AAQ20666-8). The SKC-2 gene was inserted into yeast expression vector
XX CC pPS-7 for extracellular expression of streptokinase and into pNAO
XX CC for intracellular expression in yeast. For expression in bacteria,
XX CC the SKC-2 gene was inserted in plasmid pBK33, between a trp promoter
XX CC and a T4 terminator.
XX SQ Sequence 414 AA;

Query Match 6.6%; Score 92.5; DB 13; Length 414;
Best Local Similarity 25.5%; Pred. No. 0.54;
Matches 56; Conservative 33; Mismatches 90; Indels 41; Gaps 10;

QY 52 DDLMOVIDGASRILERSLSRSRVIANALVPFHDPIGPTLTADPVVFORLRLQLYVQ 111
DB 95 ddyfveidfas---datidrn-----gkvyfadkd-gsvtlptqpv-----geffls 138
QY 112 GG---GDCPEMSVGAIAAEEVANPVSFIYVFSARAKDYHKKNELLQLLQLKQ---SQV 165
DB 139 ghvrvrpykekpigpqaksdvveytvqfclpdpddfrpglkdtklktlaigdttsge 198
QY 166 VFVLGDCGDRTHPGYLAPEEIAS--TSSGQVFDLDKQOVSEVLKWEESAIOASKVHLLS 223
DB 199 llaqaqslinkthpdytyierdsivthdnidfrtilpmdqgeftyhvkneqayeinkks 258
QY 224 ADHEE-----EGEHTWRIPDPS-LKEVTI 247
DB 259 gineeinntdlisekyyvllkgekyd-pfdrshklkfti 297

RESULT 7
AAR12892
ID AAR12892 standard; Protein; 369 AA.
AC AAR12892;
XX AAR12892;
DT 17-SEP-1991 (first entry)
XX Truncated Met streptokinase.
XX DE Fusion protein; blood clotting; coagulation; fibrinolysis;
KW

KW antithrombotic; thrombolysis; streptokinase; thrombin.
 XX
 OS Streptococcus equisimilis ATCC 9542 or ATCC 10009.
 XX

Key Location/Qualifiers
 FT Protein 2..369
 FT /label= core streptokinase
 FT /note= "AAs 16-383"
 XX

PN WO9109125-A.

XX 27-JUN-1991.

XX 07-DEC-1990; 90WO-GB01911.

XX 07-DEC-1990; 90WO-GB01911.

PR 07-DEC-1989; 89GB-0027722.

XX (BRBI-) BRIT BIO-TECHN LTD.

XX Dawson KM, Hunter MG, Czaplowski LG;

XX WPI; 1991-208151/28.

DR N-PSDB; AAQ12159.

XX Fusion protein cleavage by blood clotting enzyme - for prodn. of
 XX fractions having greater antithrombotic activity for therapy and
 XX prophylaxis.

XX Disclosure; Page 88; 115pp; English.

XX The sequence was deduced from DNA obtd. from PCR amplified chromo-
 CC somal DNA from S. equisimilis (Lancefield's Gp C.) ATCC 10009 or
 CC ATCC 9642. The gene was truncated by 15 codons at the 3' ter-
 CC minial and engineered to add a Met codon prior to codon 16 of the
 CC streptokinase CDS. The construct was used to prepare expression
 CC vector pGC606 for prodn. of the streptokinase core molecule.
 CC See also AAR12887-R12889, AAR12891-R12893, AAR12885 and AAR12522.
 XX

XX Sequence 369 AA;

Query Match 6.5%; Score 91.5; DB 12; Length 369;
 Best Local Similarity 25.3%; Pred. No. 0.57;
 Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;

QY 52 DDLMOVIDGASRIILERSLSRSRVIANVALVPFHPDIPGVTLTADPV-----VFQR 103

Db 81 ddyfevidfas---datitdrn-----gkvyfadkd-gsvtlptqpqefllsgghrvr 130

QY 104 ELRQLYVGGGDCPEMSVGAIKAAVEANPGSFYVFSARAKDYHKKNELLQLKQ- 162

Db 131 pykekpqnqa-----ksvdvetyvtqfptlppddfrpglkdtklkltaigdt 179

QY 163 --SQVFLVLTGDCDTRHPGVLAFEEIAS--TSSGVFQDLKQOVSEVLKWWESAIOASK 218

Db 180 itsqellaqaqslinkthpgygtiyerdssivthdnifrtilpmdqeftyhvkneqaye 239

QY 219 VHLLSADHEE-----EGEHTWRIPFDPDS-LKEVTI 247

Db 240 inkkslineeinntdlisekyvylkgekpyd-pfdrshlklfti 283

RESULT 8

AY84006

XX AAY84006 standard; Protein; 372 AA.

XX AAY84006;

XX 03-JUL-2000 (first entry)

XX Amino acid sequence of a mutant streptokinase.

XX

KW Streptokinase; SKC-2; plasminogen; plasmin; antigenicity;
 KW plasminogen activator complex formation; thrombolytic;
 KW myocardial infarction; pulmonary thromboembolism; thrombosis.
 XX

OS Synthetic.

OS Streptococcus equisimilis.

XX EP985729-A2.

XX 15-MAR-2000.

XX 13-AUG-1999; 99EP-0202639.

XX 14-AUG-1998; 98CU-0000119.

XX (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.

XX Torrens Madrazo IDC, Garcia Ojalvo A, De La Fuente Garcia JDJ;

PI Seralena Menendez A;

XX WPI; 2000-226041/20.

DR N-PSDB; AA299251.

XX Recombinant production of streptokinase mutants, useful as thrombolytic
 PT agents for treating myocardial infarction, that are truncated at the N-
 PT or C-terminus

XX Disclosure; Page 22-23; 54pp; English.

XX The present sequence represents a mutant protein of the streptokinase
 CC SKC-2. Streptokinase forms a complex with plasminogen, activating its
 CC conversion to plasmin. The SKC-2 gene was modified to produce the
 CC mutant protein. The obtained mutants conserve their capacity for
 CC plasminogen activator complex formation, thus having reduced
 CC antigenicity. The mutants are derived from the 1-119 gene fragment,
 CC and retain the thrombolytic activity of SKC-2 but, compared with the
 CC full-length protein, are less antigenic and retain activity better in
 CC presence of antibodies that neutralize activity of the complete
 CC protein. The mutants are thrombolytic agents which are useful for
 CC treating myocardial infarction, pulmonary thromboembolism, surgical
 CC complications and other forms of thrombosis.
 XX

XX Sequence 372 AA;

Query Match 6.5%; Score 91.5; DB 21; Length 372;
 Best Local Similarity 25.3%; Pred. No. 0.58; Indels 51; Gaps 10;

Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;

QY 52 DDLMOVIDGASRIILERSLSRSRVIANVALVPFHPDIPGVTLTADPV-----VFQR 103

Db 95 ddyfevidfas---datitdrn-----gkvyfadkd-gsvtlptqpqefllsgghrvr 144

QY 104 ELRQLYVGGGDCPEMSVGAIKAAVEANPGSFYVFSARAKDYHKKNELLQLKQ- 162

Db 145 pykekpqnqa-----ksvdvetyvtqfptlppddfrpglkdtklkltaigdt 193

QY 163 --SQVFLVLTGDCDTRHPGVLAFEEIAS--TSSGVFQDLKQOVSEVLKWWESAIOASK 218

Db 194 itsqellaqaqslinkthpgygtiyerdssivthdnifrtilpmdqeftyhvkneqaye 253

QY 219 VHLLSADHEE-----EGEHTWRIPFDPDS-LKEVTI 247

Db 254 inkkslineeinntdlisekyvylkgekpyd-pfdrshlklfti 297

RESULT 9

AY84007

XX AAY84007 standard; Protein; 384 AA.

XX AAY84007;

XX 03-JUL-2000 (first entry)

XX

XX DE Amino acid sequence of a mutant streptokinase.

XX KW Streptokinase; SKC-2; plasminogen; plasmin; antigenicity;

XX KW plasminogen activator complex formation; thrombolytic;

XX KW myocardial infarction; pulmonary thromboembolism; thrombosis.

XX OS Synthetic.

XX OS Streptococcus equisimilis.

XX PN EP985729-A2.

XX PD 15-MAR-2000.

XX PF 13-AUG-1999; 99EP-0202639.

XX PR 14-AUG-1998; 98CU-0000119.

XX PA (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.

XX PI Torrens Madrazo IDC, Garcia Ojalvo A, De La Fuente Garcia JDJ;

XX PI Seralena Menendez A;

XX DR WPI; 2000-226041/20.

XX DR N-PSDB; AAZ99252.

XX PT Recombinant production of streptokinase mutants, useful as thrombolytic

XX PT agents for treating myocardial infarction, that are truncated at the N-

XX PT of C-terminus

XX PS Disclosure; Page 24-25; 54pp; English.

XX CC The present sequence represents a mutant protein of the streptokinase

XX CC SKC-2. Streptokinase forms a complex with plasminogen, activating its

XX CC conversion to plasmin. The SKC-2 gene was modified to produce the

XX CC mutant protein. The obtained mutants conserve their capacity for

XX CC plasminogen activator complex formation, thus having reduced

XX CC antigenicity. The mutants are derived from the 1-1119 gene fragment,

XX CC and retain the thrombolytic activity of SKC-2 but, compared with the

XX CC full-length protein, are less antigenic and retain activity better in

XX CC presence of antibodies that neutralize activity of the complete

XX CC protein. The mutants are thrombolytic agents which are useful for

XX CC treating myocardial infarction, pulmonary thromboembolism, surgical

XX CC complications and other forms of thrombosis.

XX SQ Sequence 384 AA;

Query Match 6.5%; Score 91.5; DB 21; Length 384;

Best Local Similarity 25.3%; Pred. No. 0.61;

Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;

QY 52 DDLQVIGASRLERSLSRSRVIANVALPFFHDPDTPVTLTADPV-----VFQR 103

DB 95 ddyfevidfas---datitdrn-----gkyvfadkd-gsvtlptqpvqgefllghvr 144

QY 104 ELRLQYVGGGDCPMSVGAITAAVEANPGSFYVFSDDAKDYHKKNELLLQLKQ- 162

DB 145 pykekpignqa-----ksvdyeyvtqftplnpdddfpgrlktklkltlaigt 193

QY 163 --SQVFFVLTGCDGDRTHPGYLAPEEIAS--TSSGOVFDKQVSEVLKWWESAIOASK 218

DB 194 itsqellaqaqsilnkthpgytltyerdsstvtndndifrlpmdqeflyhvknreqaye 253

QY 219 VHLISADHEE-----EGEHTWRIFPDPS-LKEVTI 247

DB 254 inkkslineeintdlisekyyvikkgkpyd-pfdrshlklfti 297

RESULT 10

RAY84004

ID AAY84004 standard; Protein: 401 AA.

XX

AC AAY84004;

XX 03-JUL-2000 (first entry)

XX DE Amino acid sequence of a mutant streptokinase.

XX KW Streptokinase; SKC-2; plasminogen; plasmin; antigenicity;

XX KW plasminogen activator complex formation; thrombolytic;

XX KW myocardial infarction; pulmonary thromboembolism; thrombosis.

XX OS Synthetic.

XX OS Streptococcus equisimilis.

XX PN EP985729-A2.

XX PD 15-MAR-2000.

XX PF 13-AUG-1999; 99EP-0202639.

XX PR 14-AUG-1998; 98CU-0000119.

XX PA (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.

XX PI Torrens Madrazo IDC, Garcia Ojalvo A, De La Fuente Garcia JDJ;

XX PI Seralena Menendez A;

XX DR WPI; 2000-226041/20.

XX PT Recombinant production of streptokinase mutants, useful as thrombolytic

XX PT agents for treating myocardial infarction, that are truncated at the N-

XX PT of C-terminus

XX PS Disclosure; Page 18-20; 54pp; English.

XX CC The present sequence represents a mutant protein of the streptokinase

XX CC SKC-2. Streptokinase forms a complex with plasminogen, activating its

XX CC conversion to plasmin. The SKC-2 gene was modified to produce the

XX CC mutant protein. The obtained mutants conserve their capacity for

XX CC plasminogen activator complex formation, thus having reduced

XX CC antigenicity. The mutants are derived from the 1-1119 gene fragment,

XX CC and retain the thrombolytic activity of SKC-2 but, compared with the

XX CC full-length protein, are less antigenic and retain activity better in

XX CC presence of antibodies that neutralize activity of the complete

XX CC protein. The mutants are thrombolytic agents which are useful for

XX CC treating myocardial infarction, pulmonary thromboembolism, surgical

XX CC complications and other forms of thrombosis.

XX SQ Sequence 401 AA;

Query Match 6.5%; Score 91.5; DB 21; Length 401;

Best Local Similarity 25.3%; Pred. No. 0.65;

Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;

QY 52 DDLQVIGASRLERSLSRSRVIANVALPFFHDPDTPVTLTADPV-----VFQR 103

DB 82 ddyfevidfas---datitdrn-----gkyvfadkd-gsvtlptqpvqgefllghvr 131

QY 104 ELRLQYVGGGDCPMSVGAITAAVEANPGSFYVFSDDAKDYHKKNELLLQLKQ- 162

DB 132 pykekpignqa-----ksvdyeyvtqftplnpdddfpgrlktklkltlaigt 180

QY 163 --SQVFFVLTGCDGDRTHPGYLAPEEIAS--TSSGOVFDKQVSEVLKWWESAIOASK 218

DB 181 itsqellaqaqsilnkthpgytltyerdsstvtndndifrlpmdqeflyhvknreqaye 240

QY 219 VHLISADHEE-----EGEHTWRIFPDPS-LKEVTI 247

DB 241 inkkslineeintdlisekyyvikkgkpyd-pfdrshlklfti 284

RESULT 11

AAY84005

ID XX AAY84005 standard; Protein; 413 AA.
AC AAY84005;
XX 03-JUL-2000 (first entry)
DT
DE Amino acid sequence of a mutant streptokinase.
XX
XX Streptokinase; SKC-2; plasminogen; plasmin; antigenicity;
KW plasminogen activator complex formation; thrombolytic;
KW myocardial infarction; pulmonary thromboembolism; thrombosis.
XX
XX Synthetic.
OS Streptococcus equisimilis.
XX
XX EP985729-A2.
XX
XX 15-MAR-2000.
XX
XX 13-AUG-1999; 99EP-0202639.
XX
XX 14-AUG-1998; 98CU-0000119.
XX
XX (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.
XX Torrens Madrazo IDC, Garcia Ojalvo A, De La Fuente Garcia JDJ;
PI Seralena Menendez A;
XX
XX WPI; 2000-226041/20.
DR N-PSDB; AAZ99250.
XX
XX Recombinant production of streptokinase mutants, useful as thrombolytic
PT agents for treating myocardial infarction, that are truncated at the N-
PT of C-terminus
XX
XX Disclosure; Page 20-21; 54pp; English.
XX
XX The present sequence represents a mutant protein of the streptokinase
CC SKC-2. Streptokinase forms a complex with plasminogen, activating its
CC conversion to plasmin. The SKC-2 gene was modified to produce the
CC mutant protein. The obtained mutants conserve their capacity for
CC plasminogen activator complex formation, thus having reduced
CC antigenicity. The mutants are derived from the 1-1119 gene fragment,
CC and retain the thrombolytic activity of SKC-2 but, compared with the
CC full-length protein, are less antigenic and retain activity better in
CC presence of antibodies that neutralize activity of the complete
CC protein. The mutants are thrombolytic agents which are useful for
CC treating myocardial infarction, pulmonary thromboembolism, surgical
CC complications and other forms of thrombosis.
XX
XX Sequence 413 AA;
SQ
Query Match 6.5%; Score 91.5; DB 21; Length 413;
Best Local Similarity 25.3%; Pred. No. 0.68;
Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;
Qy 52 DDLMOVIGASRILERSLSRSRYANVALVPFHPDIPGVTLTADPV-----VFQR 103
Db 82 ddyfevidfas---datitdrn-----gkyvfadkd-gsvtlptqpvgfllsgghvr 131
Qy 104 ELRLQYVGGGDCPEMSVGAIAKAAVEVANPGSFYVFSARAKDYHKKNELLQLLQKQ- 162
Db 132 pykekipnqa-----ksvdyeyvtgftplnppddfrpglkdtklkltaigt 180
Qy 163 ---SQVFLVTGCDRTHPGYLAFFETIAS--TSSQVQLDKQVSEVLKWWESAIQASK 218
Db 181 itsqellaqaqsilnktbpgytiyerdssivthdnidfrilpmdqeftyhvkneqaye 240
Qy 219 VHLISADHEE-----EGEHTWRIPDPDS-LKEVTI 247
Db 241 inksqglneeintnlisekyyvkkgekyd-pfrdshlklftl 284

RESULT 12
AAW86143
ID AAW86143 standard; Protein; 414 AA.
XX
XX AAW86143;
AC
XX 03-MAR-1999 (first entry)
DT
XX Streptokinase (SK) protein sequence.
DE
XX Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;
KW Immunogl bulin; therapeutic; streptokinase.
XX
XX Streptococcus equisimilis.
OS
XX WO9852976-A1.
XX
XX 26-NOV-1998.
XX
XX 21-MAY-1998; 98WO-GB01473.
XX
XX 14-APR-1998; 98GB-0007751.
PR 21-MAY-1997; 97GB-0010480.
PR 31-JUL-1997; 97GB-0016197.
PR 28-NOV-1997; 97GB-0025270.
PR 02-DEC-1997; 97US-0067235.
XX
XX (BIOV-) BIOVATION LTD.
PA
XX Carr FJ;
XX
XX WPI; 1999-045301/04.
XX
XX Reducing immunogenicity of proteins - by modifying the amino acid
PT sequence of the protein to eliminate potential epitopes for T-cells
PT of a given species
XX
XX Example 6; Fig 28; 77pp; English.
XX
XX The invention relates to a method for the production of non-immunogenic
CC proteins. The method comprises determining at least part of the amino
CC acid sequence of the protein; (b) identifying in the amino acid sequence
CC one or more potential epitopes for T-cells (T-cell epitopes) of the given
CC species; and (c) modifying the amino acid sequence to eliminate at least
CC one of the T-cell epitopes identified in step (b) thereby to eliminate or
CC reduce the immunogenicity of the protein when exposed to the immune
CC system of the given species. A method of analysing a pre-existing protein
CC to predict the basis for immunogenic responses is also provided. The
CC methods can be used particularly for reducing the immunogenicity of
CC immunoglobulins or therapeutic proteins, e.g. streptokinase (SK). The
CC products can be used for diagnosis and therapy. The present sequence
CC represents the amino acid sequence of the SK protein.
XX
XX Sequence 414 AA;
SQ
Query Match 6.5%; Score 91.5; DB 20; Length 414;
Best Local Similarity 25.3%; Pred. No. 0.68;
Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;
Qy 52 DDLMOVIGASRILERSLSRSRYANVALVPFHPDIPGVTLTADPV-----VFQR 103
Db 95 ddyfevidfas---datitdrn-----gkyvfadkd-gsvtlptqpvgfllsgghvr 144
Qy 104 ELRLQYVGGGDCPEMSVGAIAKAAVEVANPGSFYVFSARAKDYHKKNELLQLLQKQ- 162
Db 145 pykekipnqa-----ksvdyeyvtgftplnppddfrpglkdtklkltaigt 193
Qy 163 ---SQVFLVTGCDRTHPGYLAFFETIAS--TSSQVQLDKQVSEVLKWWESAIQASK 218
Db 194 itsqellaqaqsilnktbpgytiyerdssivthdnidfrilpmdqeftyhvkneqaye 253

QY 219 VHLISADHEE-----EGEHTWRIPFDPDS-LKEVTI 247
 Db 254 inksglneeinntdlisekyvllkgekpyd-pfdrshklfti 297

RESULT 13

AAW86144
 ID AAW86144 standard; Protein; 414 AA.
 AC AAW86144;
 XX
 XX 03-MAR-1999 (first entry)
 XX De-Immunised streptokinase (SK) protein sequence.
 XX Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;
 KW immunogl bulin; therapeutic; streptokinase; de-immunised.
 XX Streptococcus equisimilis.
 XX WO9852976-A1.
 XX 26-NOV-1998.
 XX 21-MAY-1998; 98WO-GB01473.
 XX 14-APR-1998; 98GB-0007751.
 PR 21-MAY-1997; 97GB-0010480.
 PR 31-JUL-1997; 97GB-0016197.
 PR 28-NOV-1997; 97GB-0025270.
 PR 02-DEC-1997; 97US-0067235.
 XX (BIOV-) BIOVATION LTD.
 PA Carr FJ;
 PI WPI; 1999-045301/04.
 DR
 XX Reducing immunogenicity of proteins - by modifying the amino acid
 PT sequence of the protein to eliminate potential epitopes for T-cells
 PT of a given species

Example 6; Fig 29; 77pp; English.

The invention relates to a method for the production of non-immunogenic proteins. The method comprises determining at least part of the amino acid sequence of the protein; (b) identifying in the amino acid sequence one or more potential epitopes for T-cells (T-cell epitopes) of the given species; and (c) modifying the amino acid sequence to eliminate at least one of the T-cell epitopes identified in step (b) thereby to eliminate or reduce the immunogenicity of the protein when exposed to the immune system of the given species. A method of analysing a pre-existing protein to predict the basis for immunogenic responses is also provided. The methods can be used particularly for reducing the immunogenicity of immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The products can be used for diagnosis and therapy. The present sequence represents the amino acid sequence of a de-immunised SK protein molecule.

Sequence 414 AA;

Query Match 6.5%; Score 91.5; DB 20; Length 414;
 Best Local Similarity 25.3%; Pred. No. 0.68;
 Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;
 QY 52 DDLQVIGASRILERSLSRSRVIANVLPFHDPIGPTLTADPY-----VFQR 103
 Db 95 ddyfevdfas---datitdrn-----gkvyfadkd-gsvtlptqpvgfllghvrvr 144
 QY 104 ELRLIYVGGDCPMSVGAIKAAVEVANGPSFIYVFSDAKADYHKNNELLQLLQK- 162
 Db 145 pykekipnqa-----ksvdvevtvgtfplnpddfrpglktklkltlaigt 193

QY 163 --SQVFLVLTGDCGRTHPCYLAPEETAS--TSSGVQVQLDKQVSEVLKWFSAIQASK 218
 Db 194 itsqellaaqsilnktphgtyierdssivthndifrlpmdqeftyhvkncraye 253
 QY 219 VHLISADHEE-----EGEHTWRIPFDPDS-LKEVTI 247
 Db 254 inksglneeinntdlisekyvllkgekpyd-pfdrshklfti 297

RESULT 14

AAW01295
 ID AAW01295 standard; protein; 414 AA.
 XX
 XX AAW01295;
 XX 25-SEP-2000 (first entry)
 XX Wild type streptokinase from Streptococcus equisimilis.
 DE
 XX Immunogenicity; immunogen; T cell epitope; T-lymphocyte; drug;
 KW vaccine; carrier; diagnosis; therapy; complement; C3; C5; C9;
 KW plasminogen; streptokinase; fibrin; blood clot; thrombolysis;
 KW plasmin; myocardial infarction.
 XX Streptococcus equisimilis.
 OS WO200034317-A2.
 XX 15-JUN-2000.
 XX 08-DEC-1999; 99WO-GB04119.
 XX 08-DEC-1998; 98GB-0026925.
 PR 02-FEB-1999; 99GB-0002139.
 XX (BIOV-) BIOVATION LTD.
 PA Carr FJ, Adair FS, Hamilton AA, Carter G;
 PI WPI; 2000-423372/36.

Rendering a protein non-immunogenic or less immunogenic useful in medicine and in diagnostics involves determining the amino acid sequence of the protein, identifying and modifying potential epitopes
 Example 2; Fig 5; 42pp; English.

Proteins or their fragments can be rendered non-immunogenic or less immunogenic by identifying one or more potential T cell epitopes and modifying the sequence to eliminate at least one of these T cell epitopes to reduce the immunogenicity of the protein when exposed to the immune system of another organism. The method can be used for producing a protein with an enzymatic activity which has a beneficial therapeutic effect, a protein used to convert inactive drugs to its active form within a living organism, a vaccine, a protein used as a carrier of other molecule or a protein which binds to other molecules within or introduced within the living organism in order to alter the bio distribution of other molecules, such that the protein does not stimulate immune response in the living organism. The less immunogenic protein is useful in medicine, diagnosis and in manufacture of a therapeutic or a diagnostic agent. Streptokinase is produced by certain strains of beta-haemolytic streptococci. The protein has no inherent enzymatic activity but has considerable clinical importance owing to its ability to efficiently bind human plasminogen, potentiating its activation to plasmin and thereby promoting the dissolution of fibrin filaments in blood clots. Streptokinase is an effective thrombolytic agent in the treatment of coronary thrombosis, improving survival and preserving left ventricular function following myocardial infarction. The native protein is immunogenic and the production of neutralising antibodies in humans generally limits the protein to a single use. The new method could provide a longer therapeutic use for streptokinase. This is the sequence of the wild type streptokinase. The altered streptokinase sequence

A; Residues: 1-2014 <WI3>
A; Cross-references: EMBL:Z66566; PIDN:CAA91487.1; GSPDB:GN00028; CBSP:T25C12.3
A; Experimental source: clone T25C12
C; Genetics:
A; Gene: CBSP:T25C12.3
A; Map position: X
A; Introns: 10/2; 91/3; 136/1; 159/2; 182/3; 203/1; 225/3; 280/1; 422/3; 466/3; 559/3; 693/1; 1722/1; 1779/3; 1843/3; 1908/2; 1941/2; 2007/2

Query Match	7.8%	Score 109;	DB 2;
Best Local Similarity	23.0%	Pred. No. 1.1;	
Matches 62;	Conservative 47;	Mismatches 117;	Indels 44;
Gaps 12;			

```
Qy      26 EPTKTLSPATGDATLAFVFDVTGSWDDLMQVIDGASRIILERSLSRSR-----VIANY 79  
       ||   |   || : : : : | : | : | : | : | : | :  
Db     1381 EPPYPTSFNDKORTLVLMLETSYNMGSSIFQL----KKNIKASLDISINDDTTQGWFNF 1436
```

Qy 80 ALVPF-----HDPDIGPVLTA---DPVVFRQLRYVQG---GGDCPEMSVGAKAAV 128
||| : ||| : : ||| : : ||| : :
Dd 1437 VLYPFDSITNKDSWYPPPTISRNDDIVTALKNIETMCPGSAPCSSQCPRIVSVLSIL 1496

[illegible]

Qy 182 LAFEEIASTSSGOVFOLDKQQVSE--VLKWV-----ESAIQASKVHLLSADHEEGEHTW 234
| : | | | : : : : | : | : | : | : |
Db 1557 DALFOIISYSQGNTFTMNAVDLSKNFLTQYVPTLYSSGGTAASNGCQTDEIIIFQVEH-- 1614

```
Qy 235 RIPFPSLKVEITSLSGPGE-IEVRDPLG 263
      : | : | | : | : |
Db 1615 -----EMYEFSIDFYHPLMETIKVFDPSG 1638
```

RESULT 5
I50726

C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 20-Aug-1999
C:Accession: I50726

Proc. Natl. Acad. Sci. U.S.A. 92, 9762-9766, 1995
A; Title: Cloning and expression of the cDNA of chicken cation-independent mannose 6-phosphate phosphatase
A; Reference number: I50726; MUID:96003859

A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-2470 <2HO>

C;Superfamily: mannose 6-phosphate receptor, cation-independent; fibronectin
F;1882-1921/Domain: fibronectin type II repeat homology <2F1>

Query Match 7.6%; Score 106; DB 2; Length 2470;
Best Local Similarity 23.4%; Pred. No. 2.7;
Matches 43; Conservative 30; Mismatches 59; Indels 52; Gaps

[illegible]

Qy 73 SR-VIANYALVPHPDPTGPTLTA---DPVTFQRELRLYLVOGGDCPEMSVGAIAAA 127
:: || | | | | | | | | : : : : : | |
Db 1533 AKSSCANGAAVCTID---GPKTLNAGKLSKLTLYEDQVLKLIVYEDGDPCP----- 1579

Qy 128 VEVANPGSFIYVFS DARAKDYHKNE LLQLLQK SQVVFLTGCGDRTHPGYLAFEEI 187
:
Dd 1580 -----TDLTKT--HK-----SYFSEVKCKSDAGDDSQPVFLSFDEQ 1612

Qy 188 ASTS 191
 ||
 Db 1613 TCTS 1616

RESULT 6

C84346
hypothetical protein Vng1953c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84346
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.;
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.C.
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hobb, S.; Daniele, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe,
A.; Title: Genome sequencing of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: C84346
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1363 <STO>
A:Cross-references: GB:AE004437; NID:g10581390; PIDN:AAG20135.1; GSPDB:GN00138
A:Genetics:
A:Gene: VNG1953C

Query Match	7.5%	Score 105.5;	DB 2;	Length 1363;
Best Local Similarity	22.2%;	Pred. No. 1.3;		
Matches 58;	Conservative 33;	Mismatches 91;	Indels 79;	Caps 10;

```

QY      13  ATYTVVAAAVTSDP-----TKTLSPATGDATLAFVFDVTGSMWDDLMOVIDCASRI   64
DB      298  AASMAITADATDTPPLVVVTPSTRVQNQSDALDYAIRYN-----QPVSGSAD   347

```

QY	65	LERSLS	-----SRSRVIAN-----	YALVPFHDPDGPV	92
	:	:	:	:	:
	:	:	:	:	:
DB	348	VSMTLAAANGTENATIEDISPLVANATQRUTLDAENAPLADGVYVTLSTVWDDGSGSPV	407		

Qy	93	TLTADPVVFORELRLQYVQGSGDCPEMSVGAIKAAVEEVANPGSFI---	VYFSDARAKDYH	149
		: :		
Dd	408	TRTDPVVVIDD-----EPDVSAVSLNASTDV-SPRDTVAVNITYDDA		450

QY	150	KKNELLQLOLQKQSVFVLTCDCGTHPCYLAFAEIAS*SSGVQFQLDKQOVSVLKW	209
Db	451	-----TNATSATVHFVAARDAGNFT--AGDLNAAASA*SKDIDMAPGKQVDNLR	501

QY	210	VESAIQASKVHLLSADHEEEG	230
		: :	
Db	502	VREVTDSNYYVYVYATDERG	522

RESULT 7
T27707
hypothetical protein ZK1193.3 - *Caenorhabditis elegans*

```
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27707
C:Gelsel, C. et al. 1999. Journal of Molecular Evolution 49: 1-11.
```

A:Description: The sequence of *C. elegans* cosmid zkl193.
A:Reference number: Z20409
A:Accession: T27707

A: Molecule type: DNA
A: Residues: 1-476 <GEI>
A: Cross-references: EMBL:U41553; PIDN:AAA83292.1; CESP:2K1193.3

A:Gene: CESP:ZK1193_3
A:Introns: 22/3; 132/1; 182/3; 308/1; 358/3; 403/1

Query Match 7.5%; Score 105; DB 2; Length 476;
Best Local Similarity 22.3%; Pred. No. 0.31;
Matches 53; Conservative 49; Mismatches 100; Indels 36; Gaps 10;

Qy	38		ATLAFVFDVTGSMDDMLMQVIDGASRILE-----RSLSRSRVIANVALYVPFHD 86 : : : :
Dd	82	SAIAFAIVRSSMSKAQLNEAIEAATNTVEHYELHPIFFOKFILT---VVSNAITFSHE 141 : : : :	
Qy	87	PDIGPVTLTADPVVFQRELRLQIYV-QGGGDCPCMSYGAIKAAEVA----NPGSFITYVES 141 : : : :	
Dd	139	YDICE-----DFANSIRSLVAPPTETECDDLALAGISKTLENSAFKKYPNSPVFVES 190 : : : : :	
Qy	142	DARAK-DYHKKNELLLOLLQLQSQQVVFLTV----GDGC-DRTHPGYLAFEETASTSSGOV 195 : : : : :	
Dd	191	DGTANDFTTAGPLMEQIVNTRAQILFMITESAGSCNVDSVINIFESIRSLSQLSRGLL 250 : : : : :	
Qy	196	FOLDKQOQSEVLKWVESAIQAOSKVHLSDAHDEEGEHTWRIPF--DPSSLKEVTYSLSG 251 : : : : :	
Dd	251	IOTSLMQLSDATFSVAODLMQYDT-IILTNDLECKRAPMFQFPFFVDQSIDFLTLRASG 307 : : : : :	
	RESULT	8	
	Tl6305		
			hypothetical protein F40F4.6 - Caenorhabditis elegans
	C:Species:		Caenorhabditis elegans
	C>Date:	20-Sep-1999	#sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
	C:Accession:	Tl6305	R:Wilson, R.
			submitted to the EMBL Data Library, November 1995
	A:Description:		The sequence of C. elegans cosmid F40F4.
	A:Reference number:	Z18493	
	A:Accession:	Tl6305	
	A>Status:		preliminary; translated from GB/EMBL/DDBJ
	A:Molecule type:	DNA	
	A:Residues:	1-2214	<WILL>
	A:Cross-references:	EMBL:U04020; NID:g1065513; PID:g1065514; PIDN:AAA81430	
	C:Genetics:	G:Date:	CESP:F40F4.6
	A:Introns:	57/3; 95/1; 302/3; 323/1; 380/1; 404/1; 468/1; 507/2; 547/3; 595/3	
	Query Match	7.5%;	Score 105; DB 2; Length 2214;
	Best Local Similarity	19.4%;	Pred. No. 2.8;
	Matches	51; Conservative 46;	Mismatches 120; Indels 46; Gaps
Qy	39	TLAFFVFDVTGSMDDMLMQVIDGASRIERSLSRSRVIANVALYVPFHDPDGPTLVL----	94
Dd	1548	SLLMVVETTFRQNSDTVNIQLIAMLNKIVSANTVMPLFTNFGLVTF---DTGRTEFKED	1604
Qy	95	--TADPVVFORELRLQIYVOGGDCPCMSYGAIKAAEVA----VANPGSIYIFYVEDAKRDY	148
Dd	1605	YTSIDDLITTLTQTSSNAISTDGVCSMPYLGVLHAHLEDHDDVIAMPENSEIFLVTAGPSDL	1664
Qy	149	HKKNELLOLLQLKSQQVVFLV--TGDCG-----DRTHPGYLAFEETASTSSGOVQFQ	197
Dd	1665	GNVTWEVLFNTQAHLLHVYSKTCANCAFTDGVNNVRDMTWLGY-----GSSGNILF	1716
Qy	198	LDRQQOQSEVLKWVESAIQAOSKV-----HLLSDAHDEEGEHTWRIPFDPSSLKEVTIS	248
Dd	1717	TDPANIVNLFNFLPYLYCASVLQDPTGITNYTCS-----DGSLPWFVPVDINTFIYVT	1771
Qy	249	LSCGPGEIEVRDPLNGMSGSPPL 271	
Dd	1772	TSAEFGSLSKPDPLGAHAHSATDV 1794	

RESULT 9
T14070
peptide synthetase - Streptomyces fradiae (fragment)
C:Species: Streptomyces fradiae
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 01-Dec-2000
C:Accession: T14070
R:Hosted, T.J.; Baltz, R.H.
submitted to the EMBL Data Library, July 1997
A:Reference number: Z17868
A:Accession: T14070
A:Status: preliminary; translated from GB/EMBL/DDBJ

A: Molecule type: DNA
A: Residues: 1-1324 <HOS>
A: Cross-references: EMBL:AF016696; NID:g2738764; PID:g2738765; PIDN:AAC01566.1
C: Genetics:
A: Note: cps-1
C: Superfamily: Mycobacterium tuberculosis mbtE protein; acetate--CoA ligase homolo
C: Keywords: carrier protein; phosphantethine; phosphoprotein
F: 188-642/Domain: acetate--CoA ligase homology <ACL>
F: 659-727/Domain: acyl carrier protein homology <ACP>
F: 691/Binding site: phosphantethine (Ser) (covalent) #status predicted

```

Query Match      7.2%  Score 101.5; DB 2; Length 1324;
Best Local Similarity 20.6%; Pred. No. 2.7;
Matches 67; Conservative 38; Mismatches 123; Indels 97; Gaps 11;

QY 10 LLLATYTVAAAVTSDEPTKTLSPAT-----GDATLAFVFDVTVGSMWDDLMOV 57
   || || || || || || || || || || || || || || || || || || || || ||
Db 102 LLERALTAVHSAPATPTAAALAAATPATAARAAPRAAGCAPATIVDAFEARVRATPEAPV 161
   || || || || || || || || || || || || || || || || || || || || ||
QY 58 IDG-----ASRILERSLSRSRV---IANYALVPF-----HDP 87
   : || || || || || || || || || || || || || || || || || || || ||
Db 162 LAGGEELTYAELDARANRLARLLERGVGPESRVALTVSRNAMLPAVALGILKAGGCYVP 221
   : || || || || || || || || || || || || || || || || || || || ||
QY 88 -----DTGPTVLTADPVVFORELRQLYVQGGDCPEMSVGAIAAAVEVAN 132
   : || || || || || || || || || || || || || || || || || || || ||
Db 222 VGATLPREARAARILRETAPVCLLTDPDAEAARTRT-APTGDDRDENAPGVE---RVVL 277
   : || || || || || || || || || || || || || || || || || || || ||
QY 133 PGSFIVYFSDARAKDYHKKNELLQLLQKQSQVVFVLTGCGDRTPHPGYLAFEFIASTSS 192
   || || || || || || || || || || || || || || || || || || || ||
Db 278 TGALLAAFDPAPTDAERAGPLL-----PGHLAYLLHTSGSS 314
   || || || || || || || || || || || || || || || || || || || ||
QY 193 GOV--FOLDKQQQSEVLKWKVYESATQASKVHLLSADHHEEGEHTWRTPFPDPSLKXEVTSLS 250
   : || || || || || || || || || || || || || || || || || || || ||
Db 315 GRPKGVAVEHAQVTTALLSMAGTGVGADRLHRTVASTSES-----FDVSVDFTLVPLL 366
   : || || || || || || || || || || || || || || || || || || || ||
QY 251 GPQPEIEVRDPLGMSQ---GSPPLL 272
   : || || || || || || || || || || || || || || || || || || || ||
Db 367 TGGRIEIVENTLAVADRTGGEPSLL 391
   : || || || || || || || || || || || || || || || || || || || ||

```

RESULT 10

T48007

P-glycoprotein homolog T17J13.110 [similarity] - Arabidopsis thaliana

N:Alternate names: protein T17J13.110

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 18-Aug-2000

C:Accession: T48007

R:Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Lemcke, K.; M:

submitted to the Protein Sequence Database, February 2000

A:Reference number: Z24482

A:Accession: T48007

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1292 <RIE>

A:Cross-references: EMBL:AL138651

A:Experimental source: cultivar Columbia; BAC clone T17J13

C:Genetics:

A:Map position: 3

A:Introns: 123/3; 142/1; 200/3; 354/2; 529/3; 621/1; 904/3; 989/3; 1077/3; 1175/3

A:Note: T17J13.110

C:Superfamily: multidrug resistance protein; ATP-binding cassette homology

F:421-615/Domain: ATP-binding cassette homology <ABC1>

F:1066-1261/Domain: ATP-binding cassette homology <ABC2>

```

Query Match      6.9%; Score 96.5; DB 2; Length 1292;
Best Local Similarity 29.2%; Pred. No. 7.1;
Matches 42; Conservative 15; Mismatches 50; Indels 37; Gaps 7;

QY 24 SDEPTKLSPATGDATLAEVF-----DVTGSMDDDLMOVDIGASRI--LERSLSRRSR 74
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1034 SDESGLVDNNAVGDIELRHISFKPSRPDV--QIFQDCLSTRAGKTALVGESESGKST 1091

```

Qy 75 VIANALVPHDPDPVTL-----TADPVVQRELQLYVQG-G 113
 Db 1092 VIA--LLORFYDPDSGGITLDGVEIKTLQLKWLROQTGLVSOEVLNFETIRANIAYGKG 1149
 Qy 114 GDCPEMSVGAITKAEEVANPGSFI 137
 Db 1150 GDATETE--IVSAEELSNAHGFI 1170

RESULT 11
 major vault protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
 C:Accession: I53908
 R:Kickhoefer, V.A.; Rome, L.H.
 A:Title: The sequence of a cDNA encoding the major vault protein from Rattus norvegicus.
 A:Reference number: I53908; MUID:95129871
 A:Accession: I53908
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-895 <RES>
 A:Cross-references: EMBL:U09870; NID:g665922; PIDN:AAC52161.1; PID:g497940

Query Match 6.8%; Score 96; DB 2; Length 895;
 Best Local Similarity 21.7%; Pred. No. 4.7;
 Matches 65; Conservative 43; Mismatches 101; Indels 90; Gaps 16;

Qy 4 RAQLPLLLAT-----YTVAAAVTSDEPTKTLSPATGDAFLAFVFDVGTSMWDDLMQVI 58
 Db 52 RVLFAPMVTVPPRHVICIVANPVSRTQSSVLDITGVQRLRHA-DOEIRLAQDPPLY 110
 Qy 59 DGASRIERSRSRRIAN-----YALVPHDPDPDIPVTLTADPVVQRELQLYVQG 113
 Db 111 PG--EVLEKIDITPLQVLPNTALHLKALLDPEDKN-GDKVMAGDEWLFE----- 156
 Qy 114 GDCPEMSVGAITKAEEVANPGSFIYVSDARAKDYHKKNELLLQLQ---LKQSVVVF-L 169
 Db 157 -----GPGTYI-----PQKEVEVEITQATVIKONQALRLA 188
 Qy 170 TDCGDRTHPGYLAPEETASSTG-----QVFQDKQVSEVLKQVSAIQASK--VHLLS 223
 Db 189 RKECFDREGKGRVTGEELVRSVGAYLPVFE-----EVLVDVAVILTEKTAHLRA 241
 Qy 224 ADH-----EEGEHTW-----RIP--FDPSLKEVTISLSGPGPIEVRDPLG 263
 Db 242 LQNFRLRGLRHRTGEE--WLVTVQDTEAHVPDVVEVLGVVPIITLGPGRHYCVILDPNG 299

RESULT 12
 T28655
 receptor-like histidine kinase Bpds - Rhodococcus sp.
 C:Species: Rhodococcus sp.
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 R:Labbe, D.; Garnon, J.; Lau, P.C.
 J. Bacteriol. 179, 2772-2776, 1997
 A:Title: Characterization of the genes encoding a receptor-like histidine kinase and a
 A:Reference number: 220497; MUID:97252512
 A:Accession: T28655
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1576 <LAB>
 A:Cross-references: EMBL:U85412; NID:g1916863; PID:g1916864; PIDN:AAB52543.1
 A:Experimental source: strain M5
 C:Genetics:
 A:Gene: bpds

Query Match 6.7%; Score 93.5; DB 2; Length 1576;

Best Local Similarity 24.4%; Pred. No. 17;
 Matches 72; Conservative 37; Mismatches 111; Indels 75; Gaps 15;

Qy 1 MTPRAQLPLLLATYTVAAAVTSDEPTKTL-----SPATGDATLAFVFDVGTGS---MW 51
 Db 133 LTENGELV-LLDAGVTRVQLAANTNDQPTTSSIGADRGDPSRNNGTGGFSADLFGAGLVLL 191
 Qy 52 DDLMOVIDGASRIERS-----LSSRSRVIANVALVPHDPDPDIPGV-TLTADPVVFOR- 103
 Db 192 ESVAQANHTAS-VLRRTPQAGGDLTDLR---QFGLMGVDP-QURPVLTLLDPVVAQRY 246
 Qy 104 -----ELRQLYVQGG-GDCPEMSVGAITKAEEVANPGSFIYVSDARAKDYHKKNEL 154
 Db 247 QTAEALGELEALMATGGDTPPKGLLPRSDRAVPAAEPL-----VGRHTEL 294
 Qy 155 LQLLQKQSV-----VFVLGTGDCGDRTHPGYLAPEETASSTSSQVFOQLDKQVSEVLKW 209
 Db 295 AALISCLDGAASGTGSLCMLGDSG-----VGRSRLLEAVS- 330
 Qy 210 VESAIQASKVHLLSADHEEGEHTWRIPDPSPKKEVTISLSGPGPIE-VRDPLG 263
 Db 331 -EHAQHKSIVTLRAAAFDAQAPARPLGPFSSVFRDLAAHLRTHPSELQVRVEELG 384

RESULT 13
 T30898
 receptor-like histidine kinase - Rhodococcus erythropolis plasmid pTA421
 C:Species: Rhodococcus erythropolis
 C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
 C:Accession: T30898
 R:Arai, H.; Kosono, S.; Taguchi, K.; Maeda, M.; Fujii, F.; Chung, S.; Song, E.; Ku
 submitted to the EMBL Data Library, May 1998
 A:Description: Two sets of the biphenyl degradation genes on a linear plasmid of
 A:Reference number: 220927
 A:Accession: T30898
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1576 <ARA>
 A:Cross-references: EMBL:AB014348; NID:d1241341; PID:d1035241; PIDN:BAA34261.1
 A:Experimental source: plasmid TA421
 C:Genetics:
 A:Gene: bpds
 A:Genome: plasmid pTA421

Query Match 6.7%; Score 93.5; DB 2; Length 1576;
 Best Local Similarity 24.4%; Pred. No. 17;
 Matches 72; Conservative 37; Mismatches 111; Indels 75; Gaps 15;

Qy 1 MTPRAQLPLLLATYTVAAAVTSDEPTKTL-----SPATGDATLAFVFDVGTGS---MW 51
 Db 133 LTENGELV-LLDAGVTRVQLAANTNDQPTTSSIGADRGDPSRNNGTGGFSADLFGAGLVLL 191
 Qy 52 DDLMOVIDGASRIERS-----LSSRSRVIANVALVPHDPDPDIPGV-TLTADPVVFOR- 103
 Db 192 ESVAQANHTAS-VLRRTPQAGGDLTDLR---QFGLMGVDP-QURPVLTLLDPVVAQRY 246
 Qy 104 -----ELRQLYVQGG-GDCPEMSVGAITKAEEVANPGSFIYVSDARAKDYHKKNEL 154
 Db 247 QTAEALGELEALMATGGDTPPKGLLPRSDRAVPAAEPL-----VGRHTEL 294
 Qy 155 LQLLQKQSV-----VFVLGTGDCGDRTHPGYLAPEETASSTSSQVFOQLDKQVSEVLKW 209
 Db 295 AALISCLDGAASGTGSLCMLGDSG-----VGRSRLLEAVS- 330
 Qy 210 VESAIQASKVHLLSADHEEGEHTWRIPDPSPKKEVTISLSGPGPIE-VRDPLG 263
 Db 331 -EHAQHKSIVTLRAAAFDAQAPARPLGPFSSVFRDLAAHLRTHPSELQVRVEELG 384

RESULT 14
 T19941
 hypothetical protein C44H4.1 - Caenorhabditis elegans

C/Species: *Caenorhabditis elegans*
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T19941
R:Smye, R.
submitted to the EMBL Data Library, August 1996
A/Reference number: Z19200
A/Accession: T19941
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-458 <WIL>
A/Cross-references: EMBL:Z79598; PIDN:CA801867.1; GSPOB:GN00028; CESP:C44H4
A/Experimental source: clone C44H4
C/Genetics:
A/Gene: CESP:C44H4.1
A/Map position: X
A/Introns: 24/3; 157/1; 257/3; 372/3; 425/1

```

Query Match      6.6%; Score 93; DB 2; Length 458;
Best Local Similarity 19.4%; Pred. No. 3.2;
Matches 56; Conservative 51; Mismatches 99; Indels 82; Gaps 14;

QY 11 LLATYTVVAARVTSDEPTKTL-SPATGDATLAFVFDVL--GSMWDDLQWVIDGASRLLE 66
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 LLLFNLLIAGNVYSQCPTLQLOEECTCTSTREAVSINCGGSLDAVLESLSNPQAIID 62
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 67 RSLSSRSRV-----IANYALVPFHDPDIPGVT 93
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 SLTISNTPIEKVLVFNHDKISYLLFEOMPQYAFQFGQTKLFLRNNGLSRSH-----PNT 117
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 94 LTADPVVVFQRELQLYVQGG--GDCPEMSYGAIKAAVEVANPGSFYVFSARAKDYHKK 151
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 FTGN---LENSLELEIRGNVIDGIPQSGVSYLKQLKILSLPDNLIEVVDQNAFLSVHR 174
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 152 NELLQLQLQKOSQVVFVLITGDCDGRTHP-GYLAPEEETASTSSGGVFDLDQQQSEV---- 206
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 DSLLK-LDLSANNLTAI-----HPTGLLGLNLSQLS-----LDKNLLSEIPSOA 218
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 207 LKWVES-----AIQSKVHLLLSADHEEGEHTWRIPDPSLKEVTISLS 250
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 219 LENIPSLSDLSLGNRIHTLSRN-----SIPL-PNEKLSLSLEVN 256
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 15
A75207
amytopullulanase PAB0122 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A75207
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A:Reference number: A75001
A:Accession: A75207
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1362 <RAW>
A:Cross-references: GB:AJ248283; GB:AL056836; NID:g5457433; PIDN:CAB49104.1; PID:e151499
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: apu; PAB0122

```

Query Match	6.6%	Score 93;	DB 2;	Length 1362;
Best Local Similarity	21.5%	Pred. No. 15;		
Matches 65;	Conservative 42;	Mismatches 105;	Indels 90;	Gaps 14;

QY	17	VVAAAVTSDEPTK----	TLSP-ATGDATAI	FEVDVTGSMWDDL	MQVDDGASRI	LSLS	71
		: : : :	: : : :	: : : :	: : : :	: : : :	
Db	622	IVRALVGLPEGVKNSS	SFLAG----	IEVYFD-----	DEGLHFVVL	TNRSEI	672
		: : : :	: : : :	: : : :	: : : :	: : : :	
QY	72	RSRVIAN-----	YALVPFHDPD	IGPVT	LTADPV-----	VFQRELQLX	112
		: : : :	: : : :	: : : :	: : : :	: : : :	

```

Db 673 PEKIIGNTFTVLQKKPEEFRYSEVPFSDKSVGLLTIHTITVKGGEVFVKATSYDNYKKV 7332
QY 113 GGDCEPMSVGAIKAAVEVANPGSFI-----YVFS DARAKDYHKKNELLQLQLK-----Q 1622
Db 733 G----EVKYNAINGYEVVVPFDYIEPDSFYFAVSTINDNGSLEIITTPIHUKLPKEIE 788
QY 163 SQVFLVLTGDCGDTRTHGYLAFEEIASTSSGOVFLQDKQVSEVLKWSAIOASQVHLL 222
Db 789 GTLITEIKDIEGDDHGPNGVTY-----ATDKVF-----VEHHLLDLKVRLL 829
QY 223 SADH:-----BEEGEHTWRIPDPSLKEVTLSL-----SGGPEIEVR 259
Db 830 ERPSYVFEFYFKELGDNPNWAPYGLSQIMEVYLDYKEGNTSAIKMPDPGPGSNVDL- 888
QY 260 DP 261
Db 889 DP 890

Search completed: March 7, 2002, 22:06:20
Job time: 3533 sec

```

Search completed: March 7, 2002, 22:06:20
Job time: 3533 sec

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	96	6.8	885	1	MVP_RAT	Q62667 rattus norv
2	93.5	6.7	677	1	CHLD_SYNP7	O07345 synechococ
3	92.5	6.6	1387	1	PUR4_ARATH	Q9m8d3 arabidopsi
4	92	6.6	686	1	BCHD_HELMO	Q9zge6 heliobacill
5	91.5	6.5	806	1	ITB7_MOUSE	P26011 mus musculu
4	90.5	6.5	693	1	TGL3_MOUSE	Q08189 mus musculu
7	89.5	6.4	643	1	PDA4_RAT	P38659 rattus norv
8	88.5	6.3	440	1	STRP_STRSP	P10519 streptococc
9	88.5	6.3	515	1	PDL_ASPOR	Q00248 aspergillus
10	88	6.3	896	1	MVP_HUMAN	Q14764 homo sapien
11	88	6.3	1098	1	PGDR_MOUSE	P05622 mus musculu
12	87.5	6.2	158	1	GRE4_RHILY	O68546 rhizobium l
13	87	6.2	758	1	CHLD_TOBAC	O24133 nicotiana t
14	86.5	6.2	440	1	STRP_STREQ	P00779 streptococc
15	86	6.1	403	1	PTEN_HUMAN	O00633 homo sapien
16	86	6.1	403	1	PTEN_MOUSE	O08586 mus musculu
17	85.5	6.1	754	1	CHLD_PPA	O22437 pisum sativ
18	85.5	6.1	992	1	PHS2_DICDI	P34114 dictyostell
19	85	6.1	575	1	HMD1_HEVBR	P29057 hevea bras
20	85	6.0	1433	1	VGLM_BUNYW	P04505 bunyamwera
21	84.5	6.0	2128	1	SPCB_MOUSE	P15508 mus musculu
22	84	6.0	759	1	CHLD_ARATH	O9sjel arabidopsi
23	83.5	6.0	695	1	AMPH_HUMAN	P34418 homo sapien
24	83	5.9	507	1	ATPA_CHLRE	P28526 chlamydomon
25	82.5	5.9	158	1	GRE4_RHIME	P56894 rhizobium m
26	82.5	5.9	344	1	YK58_ARCFU	O28221 archaeglob
27	82.5	5.9	1260	1	SPOE_SCHPO	Q10093 schizosacch
28	82.5	5.9	2137	1	SPCB_HUMAN	P11277 homo sapien
29	82.5	5.9	3649	1	ACVS_NOCLA	P27743 nocardia la
30	82	5.8	506	1	ATPA_CHLVU	P56294 chlorella v
31	81.5	5.8	412	1	AK_PSEAE	O69077 pseudomonas
32	81.5	5.8	913	1	CAD4_MOUSE	P39038 mus musculu
33	81.5	5.8	1106	1	PGDR_HUMAN	P09619 homo sapien

Query Match	6.8;	Score 96;	DB 1;	Length 885;
Best Local Similarity	21.78;	Pred. No. 2.8;		
Matches 65;	Conservative 43;	Mismatches 101;	Indels 90;	Gaps 16;
QY	4	RAQLPLLLAT-----YTWAAAVTSDEPTKTLSPATGATLAFVDFVTGSMDDLMQVI	58	
DB	42	RVLFAPVMVAVPPRHYCIIVANPVSRRDQSSVFDITQVRLRHA-DGEIRLAQDPFFLY	100	
QY	59	DGASRILERSLSSRSRVIAN-----YALVPFHPDIPGIVTITADPVVFQRELRLQYVGG	113	
DB	101	PG--RVLEKSDITPLOVLPNTALHLKALLDEEDKN-GDKVAGDSWLF-----	146	

QY 114 GDCPEMSYGAIKAAEVANPGSFYTFVSADAKDYHKKNELLOLLOLQ---LKQSQVVFV-L 169
 Db 147 -----GPGTYI-----PQREVEVEIIQATVIKQNALRLRA 178
 QY 170 TGDGCDRTHPGYLAFAEIASTSSG-----QVFQDKQVSEVILKWKVESAIQASK--VHLLS 223
 Db 179 RKECFDREGKRVGTGEEMVRSVGYLPAVFE-----EVLDLVDVAIVLTKETALHLRA 231
 QY 224 ADH-----EEEGHTW-----RIP--FDPISLKEVTISLSPGPEIEVRDPLG 263
 Db 232 LONFRDLRGVLRHTGEE--WLVTVDTEAHVPDYVEEVLGVVPIITLGRHYCVILDPMG 289

RESULT 2

CHLD_SYN7
 ID CHLD_SYN7 STANDARD; PRT; 677 AA.
 AC O07345;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MAGNESIUM-CHELATASE SUBUNIT CHLD (MG-PROTOPORPHYRIN IX CHELATASE) (MG-CHELATASE SUBUNIT D).
 GN CHLD.
 OS *Synechococcus* sp. (strain PCC 7942) (Anacystis nidulans R2).
 OC Bacteria; Cyanobacteria; Chroococcales; *Synechococcus*.
 OX NCBI_TaxID=1140;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Masuda T., Nakayama M., Ohta H., Takayama K.-I.;
 RT "Cloning and sequencing of a chld gene encoding a subunit of
 RT magnesium-chelatase from the cyanobacterium *Synechococcus* sp. PCC
 RT 7942.";
 RL (In) Plant Gene Register PGR97-091.
 CC - FUNCTION: INVOLVED IN CHLOROPLAST PIGMENT BIOSYNTHESIS; INTRODUCES
 CC A MAGNESIUM ION INTO PROTOPORPHYRIN IX TO YIELD MG-
 CC PROTOPORPHYRIN IX.
 CC - PATHWAY: CHLOROXYLL BIOSYNTHESIS.
 CC - SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNITS D/I FAMILY.
 CC - SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB0031135; BAA20346.1;
 DR InterPro: IPR000523; MG_chelatase_chiI.
 DR Pfam: PF01078; Mg_chelatase; 1.
 DR SMART: SM00327; VWFA; 1.
 DR PROSITE: PS50234; VWFA; 1.
 KW Photosynthesis; Chlorophyll biosynthesis.
 FT DOMAIN 481 676
 FT DOMAIN 329 383
 FT SEQUENCE 677 AA; 73367 MW; 7E2988E096DF5506 CRC64;

Query Match 6.78; Score 93.5; DB 1; Length 677;
 Best Local Similarity 24.18; Pred. No. 3.2;
 Matches 52; Conservative 98; Indels 29; Gaps 9;

QY 24 SDEPTKTLSPATGATLAFVDVTGSMWDDLMQVIDCA-SKILERSLSRSRVNTANVLY 82
 Db 467 ADIRSKQLVRKAG-ALVIFLVDAGSGMALNMQSAKGAIVRLTTEAYENRDQV----ALI 521
 QY 83 PFHPDPIGPVTL--TADPVVVFQRELRLQYVQGGDCPEMSYGAIKAAEVANPGSF---- 136
 Db 522 PFRG-EQAEVLLPPTRSITTAARKRLKMKPCGGGLPLAHLGTQAVRVCTNAQSGDITGVV 580

QY 137 IYVESDARA-----KDYHK---KNELQLQLKQSQVVFVLTGCDGDRTHPGY 181
 Db 581 IVALTDGRGNIPLARSIGQPMEEGEKPDLEELDLIAKRNLGSLQMLLYDIT-PRKPVGA 639
 QY 182 LAEEIASTSSGQVFQDKQVSEVILKWKVESAIQAS 217
 Db 640 GFGEKLANAAGGQYHLPKVSQDAIAAMAQSAALRAT 675

RESULT 3

PUR4_ARATH
 ID PUR4_ARATH STANDARD; PRT; 1387 AA.
 AC Q9M8D3;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE PROBABLE PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE, CHLOROPLAST
 DE PRECURSOR (EC 6.3.5.3) (FGAM SYNTHASE) (FGAMS) (FORMYLGLYCINAMIDE
 DE RIBOTIDE AMIDOTRANSFERASE) (FGARAT) (FORMYLGLYCINAMIDE RIBOTIDE
 DE SYNTHETASE).
 GN ATIG74260 OR FIO17.7.
 OS *Arabidopsis thaliana* (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; *Arabidopsis*.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=2016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.B., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pail G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant *Arabidopsis*
 RT *thaliana*.";
 RL Nature 408:816-820(2000).
 CC - CATALYTIC ACTIVITY: ATP + 5'-PHOSPHORIBOSYLFORMYLGLYCINAMIDE +
 CC L-GLUTAMINE + H(2)O = ADP + ORTHOPHOSPHATE + 5'-PHOSPHORIBOSYL-
 CC FORMYLGLYCINAMIDINE + L-GLUTAMATE.
 CC - PATHWAY: DE NOVO PURINE BIOSYNTHESIS; FOURTH STEP.
 CC - SUBCELLULAR LOCATION: CHLOROPLAST (PROBABLE)
 CC - SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE FGAMS
 CC FAMILY.
 CC - SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO TYPE-1 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AC020579; AAG52403.1;
 DR InterPro: IPR000728; AIRS_related.
 DR Pfam: PF00586; AIRS; 1.
 KW Purine biosynthesis; Ligase; ATP-binding; Glutamine amidotransferase;
 KW Chloroplast; Transit peptide.
 FT TRANSIT 1 ?
 FT CHLOROPLAST (POTENTIAL).

[illegible]

DE (TGASE E3) (TRANSGLUTAMINASE 3).
GN TGM3 OR TGASE3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/C; TISSUE=Epidermis;
RX MEDLINE=93286109; PubMed=8099584;
RA Kim I.-G., Gorman J.J., Park S.-C., Chung S.-I., Steinert P.M.;
RT "The deduced sequence of the novel protransglutaminase E (TGase), of human and mouse."
RL J. Biol. Chem. 268:12682-12690(1993).
CC -!- FUNCTION: CATALYZES THE CROSS-LINKING OF PROTEINS AND THE CONJUGATION OF POLYAMINES TO PROTEINS. IT IS RESPONSIBLE FOR THE LATER STAGES OF CELL ENVELOPE FORMATION IN THE EPIDERMIS AND THE HAIR FOLLICLE.
CC -!- CATALYTIC ACTIVITY: PROTEIN GLUTAMINE + ALKYLAMINE = PROTEIN N(5)-ALKYLGLUTAMINE + NH(3).
CC -!- COFACTOR: CALCIUM IS REQUIRED FOR TRANSGLUTAMINASE ACTIVITY.
CC -!- SUBUNIT: THIS ENZYME CONSISTS OF TWO POLYPEPTIDE CHAINS, WHICH ARE SYNTHESIZED AS A PRECURSOR FORM OF A SINGLE POLYPEPTIDE.
CC -!- SIMILARITY: BELONGS TO THE TRANSGLUTAMINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC
CC EMBL; L10385; AAA40421.1; .
DR HSSP; P00488; 1GGT
DR MGD; MGI:98732; Tgm3.
DR InterPro; IPR002931; Transglut_core.
DR InterPro; IPR001102; Transglutmmse.
DR Pfam; PF00927; Transglutamin-C; 1.
DR Pfam; PF00868; Transglutamin-N; 1.
DR Pfam; PF01841; Transglut_core; 1.
DR SMART; SM00460; TGC; 1.
DR PROSITE; PS00547; TRANSGLUTAMINASES; 1.
KW Transferase; Acyltransferase; Calcium-binding; zymogen.
FT CHAIN 1 470 50 KDA NONCATALYTIC CHAIN.
FT CHAIN 471 693 27 KDA CATALYTIC CHAIN.
FT ACT_SITE 273 273 BY SIMILARITY.
SQ SEQUENCE 693 AA; 77324 MW; 8B3F0601735D4E84 CRC64;

Query Match 6.5%; Score 90.5; DB 1; Length 693;
Best Local Similarity 21.8%; Pred. No. 5.9;
Matches 60; Conservative 29; Mismatches 71; Indels 115; Gaps 14;

QY 48 GSWDDLMQVIGASRILERSLSRSRVIAN- 79
DB 271 GOCV-----VFAGTLNTVLRCLGVRSVITFNESAHTDRNLSDVYVYDAMGNPLEKGS 325
QY 80 ALVPEH-----DPDIPGV-----TLTADPVVQRELRLQLYVGGDCCPMSVGA 126
DB 326 SWNFHVWNEGFVTRDLPYTGNGVQLDAP-----QERSQGVFQCGPASVNAIKA 377
QY 127 AVEANPGSFIYVFSARAKD-----YHKNNELLQLLQKQSVFVLTCDCGDRTHP--G 180
DB 378 G-DVDRNEDMIFAEVNADRTIWIYNNRN-----TQKNSV-----DRHSIGK 421
QY 181 YLAFIEAISTSGQVFLDKQVSEVLKWKVESIAQSKVHLSAD-----HE 227
DB 422 YISTRKAVGNS-----RMDVTDKYKYPEGSEERVOHQALDKLPNASFGATSSRN 473
QY 228 EGEHTWRIPFDPSSL-----KEVTISL 249
DB 474 PEGEDK-----EPSSIGKFKVTGILAVKGVEVSL 503

RESULT 7
PDA4_RAT PDA4_RAT STANDARD; PRT; 643 AA.
ID AC P38659;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN DISULFIDE ISOMERASE A4 PRECURSOR (EC 5.3.4.1) (PROTEIN ERP-72)
DE (ERP72) (CALCIUM-BINDING PROTEIN 2) (CABP2).
GN PDIA4 OR CABP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Liver;
RX MEDLINE=93238767; PubMed=8477750;
RA Van P.N., Rupp K., Lampen A., Soeling H.-D.;
RT "Cabp2 is a rat homolog of ERP72 with protein disulfide isomerase activity."
RL Eur. J. Biochem. 213:789-795(1993).
CC -!- CATALYTIC ACTIVITY: REARRANGEMENT OF BOTH INTRACHAIN & INTERCHAIN DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES.
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -!- INDUCTION: UPON GLUCOSE STARVATION, AS WELL AS TREATMENT WITH TUNICAMYCIN.
CC -!- PTM: O-GLYCOSYLATED.
CC -!- SIMILARITY: CONTAINS 3 THIOREDOXIN DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC
CC EMBL; M86870; AAA19217.1; .
DR HSSP; P07237; IMEK.
DR MEROPS; C17.002; .
DR InterPro; IPR000886; ER target.
DR InterPro; IPR000063; Thioered.
DR Pfam; PF00085; thioered; 3.
DR PRINTS; PR00421; THIOREDODXIN.
DR PROSITE; PS00194; ER_TARGET; 1.
DR PROSITE; PS00194; THIOREDODXIN; 3.
KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal; Glycoprotein; Calcium-binding
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 643 PROTEIN DISULFIDE ISOMERASE A4.
FT DISULFID 89 92 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 204 207 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 553 556 REDOX-ACTIVE (BY SIMILARITY).
FT SITE 640 643 PREVENT SECRETION FROM ER.
SQ SEQUENCE 643 AA; 72806 MW; 85611B66F3A4DE41 CRC64;

Query Match 6.4%; Score 89.5; DB 1; Length 643;
Best Local Similarity 22.0%; Pred. No. 6.5;
Matches 53; Conservative 35; Mismatches 76; Indels 77; Gaps 11;

QY 15 YTVVAAVTSDEPTKLS--PATGATLAFVDFVTGSMWDDL----QVIGASRILERSL 69
DB 99 YEKIASTLKNDPPIAVAKIDATASMLASKFDVSGYPTIKLKGGQAVDVGSRQTGEI 158
QY 70 SRSRVIANALVPHDPDGP---VTLTA-----DPVVFQBE--LRQLYVGGGDCPEM 119
DB 159 VAKVREVS-----QPDWTTPPPPEVTTLTKENFDVYVNNADIILVEFYAPWCGHCKKL 210
QY 120 SVGAIKAAVEVANPG-----SFIYVFSARAKDYHKNNEL 154

```
Db 211 APEYKAAKELSKRPPILAKVDATQTLAKRFDVSGYPTLKIFKRGKRPFDYNGPREK 270
QY 155 L-----QILLOLKOSQVVFVLTCG-----CGDRTHPGYLAFEEETAS 189
Db 271 YGIVDYMVEQSGPPSKEITLTKQVO-EFLKGDGDDVVILGVFGVGVD---PGYLOQYODAA 326
QY 190 T 190
Db 327 T 327
```

```
RESULT 8
STRP_STRP STANDARD; PRT; 440 AA.
AC P10519;
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE STREPTOKINASE G PRECURSOR.
GN SKG.
OS Streptococcus sp. (strain 19909).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89160265; PubMed=2922269;
RA Walter F., Siegel M., Walke H.;
RT "Nucleotide sequence of the streptokinase gene from a group-G
RT Streptococcus.";
RL Nucleic Acids Res. 17:1262-1262(1989).
CC -!- FUNCTION: THIS PROTEIN IS NOT A PROTEASE, BUT IT ACTIVATES
CC PLASMINOGEN BY COMPLEXING WITH IT. AS POTENTIAL VIRULENCE FACTOR,
CC IT IS THOUGHT TO PREVENT THE FORMATION OF EFFECTIVE FIBRIN
CC BARRIERS AROUND THE SITE OF INFECTION, THEREBY CONTRIBUTING TO THE
CC INVASIVENESS OF THE CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13400; CAA31766.1; -.
DR PIR; S02723; S02723.
KW Plasminogen activation; Signal; Virulence.
FT SIGNAL 1 26
FT CHAIN 27 440 STREPTOKINASE G.
SQ SEQUENCE 440 AA; 50199 MW; 5521F8825FE1B6EA CRC64;
```

```
Query Match 6.3%; Score 88.5; DB 1; Length 440;
Best Local Similarity 25.3%; Pred. No. 4.7;
Matches 57; Conservative 30; Mismatches 87; Indels 51; Gaps 10;

QY 52 DDLMOVIDGASRLRSLSRSRVIANALVPFHPDPIDIGPVTLTADPV-----VFOR 103
Db 121 DDYFEVIDFAS---DATITDRN-----CKVVFADKD-GSVTLPIQPQVEFLKGHVRV 170
QY 104 ELRQLYVQGGDCPMSVGAIAKAAVEANPGSFIYVFSARAKYHKKNELLQLQLKQ- 162
Db 171 PYKEKPVQQA-----KSDVVEYTVQFTPLNPDDDFRPALKDKTKLKTLAGT 219
QY 163 --SOVVFVLTGCGDRTHPGYLAFEEIAS--TSSGOVFOLDKQVSEVLKWKVESATOASK 218
Db 220 ITSQELLAQAQILNKNHPGIIYERDSSIVTHDNDIFRTILPMDQEPFTHVKNREQAYR 279
QY 219 VHLLSADHEE-----EGEHTWRIPDPS-LKEVTI 247
Db 280 INKSKSLNEEINNTDLISEKYYVLLKKGKPYD-PFDRSHLKLFTI 323
```

```
RESULT 9
PDL_ASPOP
ID PDL_ASPOP STANDARD; PRT; 515 AA.
AC Q00248;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROTEIN DISULFIDE ISOMERASE PRECURSOR (EC 5.3.4.1) (PDI).
GN PDI.A.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIB 40;
RA Lee B., Yamada O., Kitamoto K., Takahashi K.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
CC DISULFIDE BONDS. MAY BE INVOLVED IN GLYCOSYLATION, PROLYL
CC HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: REARRANGEMENT OF BOTH INTRACHAIN & INTERCHAIN
CC DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES.
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D85900; BAA12913.1; -.
DR HSP; P07237; IMBK.
DR InterPro; IPR000886; ER_target.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; ThioRed; 3.
DR PRINTS; PR00421; THIOREDOXIN.
DR PROSITE; PS00014; ER_TARGET; 1.
DR PROSITE; PS00194; THIOREDOXIN; 2.
KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 515 PROTEIN DISULFIDE ISOMERASE.
FT DISULFID 58 61 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 393 396 REDOX-ACTIVE (BY SIMILARITY).
FT SITE 512 515 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 515 AA; 56458 MW; A155B38E20E37EAC CRC64;
```

```
Query Match 6.3%; Score 88.5; DB 1; Length 515;
Best Local Similarity 18.5%; Pred. No. 5.8;
Matches 63; Conservative 52; Mismatches 109; Indels 117; Gaps 15;

QY 7 LLPLLLATYVVAAVTSDEPTKTLSPATGATLAFVDFVTGSMWDDL----- 55
Db 8 ILSLLGASAVASAAATAEAAPSDDVVS-LTGDTTFETVKE-----HDLVLAEEFFAPWCGH 60
QY 56 -----QVLDGASRLRSLS-----SRSRVIANY-----ALVPFHD- 86
Db 61 CKALAPKYEQAATELKEKNIPLVKVDCTEEALCRDQGVGYPTLKIFGLDAVPYQGA 120
QY 87 -----PDIGPVT-----LTADPV-----FQREL 106
Db 121 RQTEAIVSMVKQSLPAVSPVTPENLEEIKTMKIVVICIASDDQANDIFTTAEOR 180
QY 107 QLYVQGGDCPMSVGAIAKAAVEANPGSFIYVFSARAKYH---KNELLQLQLKQS 163
Db 181 DNYLFAA-----TSDASTAKAEGVKQPSIVLYKDFEKKATYDGBTEODALLSWKVTAST 235
```

```

Query Match      6.3%; Score 88; DB 1; Length 896;
Best Local Similarity 21.0%; Pred. No. 14;
Matches 65; Conservative 43; Mismatches 90; Indels 112; Gaps

QY  4 RAQLPLLLAT----YTVVAAVNTSDEPKTTLSPATGDTATLAFVDDVTGS----- 49
      |         | : : : | : : | : : | : : | : : | : : | : : | : : |
Db  42 RVLFAPMRWTVVPPRHCTVA-----NPNVSRDAQGLVLFDDTVQVRLRHADLE 89
      |         | : : : | : : | : : | : : | : : | : : | : : | : : |
QY  50 --MWDDLQMQLVDGASRIILERSLSRSRVIAN-----YALVPFHDPDITGPVTLTADPVVQ 102

```

DR Pfam: PF00047; ig: 4.
DR Pfam: PF00069; pkinase; 2.
DR SMART: SM00408; IGC2; 1.
DR SMART: SM00410; IG_Like; 2.
DR SMART: SM00219; Ty_KC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00240; RECEPTOR_TYR_KIN_III; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
KW Tyrosine-protein kinase: Receptor; Transmembrane; Glycoprotein;
KW Transferase; Phosphorylation; ATP-binding; Immunoglobulin domain;
KW Signal.
FT SIGNAL 1 31
FT CHAIN 32 1098
FT FT
FT FT
FT DOMAIN 32 530
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 531 555
FT POTENTIAL.
FT DOMAIN 556 1098
FT CYTOPLASMIC (POTENTIAL).
FT IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 128 209
FT IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 214 310
FT IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 415 527
FT IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 599 613
FT PROTEIN KINASE.
FT NP_BIND 605 631
FT BINDING 632 633
FT ACT_SITE 825 825
FT DISULFID 53 99
FT DISULFID 148 189
FT DISULFID 234 290
FT DISULFID 435 507
FT MOD_RES 750 750
FT MOD_RES 856 856
FT CARBOHYD 44 44
FT CARBOHYD 88 88
FT CARBOHYD 102 102
FT CARBOHYD 214 214
FT CARBOHYD 291 291
FT CARBOHYD 306 306
FT CARBOHYD 353 353
FT CARBOHYD 370 370
FT CARBOHYD 444 444
FT CARBOHYD 467 467
FT CARBOHYD 478 478
SQ SEQUENCE 1098 AA; 122805 MW; 8D391CAFAC3FC31D CRC64;

Query Match 6.3%; Score 88; DB 1; Length 1098;
Best Local Similarity 18.7%; Pred. No. 18;
Matches 56; Conservative 45; Mismatches 107; Indels 92; Gaps 12;

QY 20 AAVTSDEPTKTLSPATGDAITLAFVFDVTGSMWDDLMQVIDGASRLERSLSRSRVIAN 79
DB 788 ATLINDSPVLSYTDLVG-----FSYQVANGM--DFL-----ASKNCVHRDLAARNVLICEG 836
QY 80 ALVPEHPDPDIG-----PV-----TLTADPVVFQRELRLQLY 109
DB 837 KLVKICDFGLARDMRSNYSISKSTYLPKWWAPESIFNSLYTTLSDVWSFGILLWEIF 896
QY 110 VQGGDGPENSV-----GAIKAAVEVAPGSIYVFSARAKDYHKKNELQLQLKQSQ 164
DB 897 TLGGTPTPELPMNDQFYNAIKRGVMAQPA-----HASDEYEINQ----- 937
QY 165 VVFLTGCGD---RTHPGYLAFFIEIASTSSGVQVFLDKQOVSEVLKWSAIAQSKVHL 221
DB 938 -----KCWEKEPFRPPESQLVLLERLLGEGYKKKYQVD-----EEFLRSHPAI 984
QY 222 LSADHEEGEHTWRIPD-----PSLKEVTISLGGPIEVRDPLGMSQSGPPL 271
DB 985 LRQAPPGIHSURKSPDLTSSVLYTAVQPNESDNDYTIPLDPKPDVADE-GLPEGSPSL 1043

RESULT 12
GREA_RHLV

ID GREA_RHLV STANDARD; PRT; 158 AA.
AC O68546;
DC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTION ELONGATION FACTOR GRE A (TRANSCRIPT CLEAVAGE FACTOR GRE A).
GN GREA.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3855;
RX MEDLINE=98434549; PubMed=9756877;
RA Kadmas J.L., Allaway D.; Studholme R.E., Sullivan J.T., Ronson C.W., Poole P.S., Raetz C.R.H.;
RT "Cloning and overexpression of glycosyltransferases that generate the lipopolysaccharide core of Rhizobium leguminosarum.";
RL J. Biol. Chem. 273:26432-26440(1998).
CC -1- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION ELONGATION PAST TEMPLATE-ENCODED ARRESTING SITES. THE ARRESTING SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY CLEAVAGE FACTORS SUCH AS GRE A OR GREB ALLOWS THE RESUMPTION OF ELONGATION FROM THE NEW 3'TERMINUS. GRE A RELEASES SEQUENCES OF 2 TO 3 NUCLEOTIDES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GRE A/GREB FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF050103; AAC05214.1; .
CC HSP: P21346; IGRJ.
CC InterPro: IPR001437; GreA_Greb.
CC Pfam: PF01272; GreA_Greb; 1.
CC ProDom: PD004918; GreA_Greb; 1.
CC PROSITE: PS00829; GREAB_1; 1.
CC PROSITE: PS00830; GREAB_2; 1.
KW Transcription regulation; DNA-binding; Coiled coil.
SQ SEQUENCE 158 AA; 17444 MW; FACA215D8CAF1605 CRC64;

Query Match 6.2%; Score 87.5; DB 1; Length 158;
Best Local Similarity 24.3%; Pred. No. 1.4;
Matches 36; Conservative 30; Mismatches 55; Indels 27; Gaps 7;

QY 133 PGSIYVFSARAKDYHKKNELQLQLKQSQVFLTGCGDTRTHPGYLAFFIEIASTSS 192
DB 9 PGGEVFKQEELRWQRQEPRIIEAIEARAH-----GDLSENAE--YHAKEAQSHNE 60
QY 193 GOVFOLD-----KQOVSEVLKWSAIAQ-ASKVHLLSADHEEGEHTWRIPD--PSLKEV 245
DB 61 GRISLEDLTARAEVIDLTKMSGDKIKFGAKVLIDEDETEB--EKTQIVGDQADYKAG 118
QY 246 TISLGGP-----GPEIEVRDPLG 263
DB 119 RISISSPIARALICKEVGDSIEVNPAGG 146

RESULT 13
CHLD_TOBAC
ID CHLD_TOBAC STANDARD; PRT; 758 AA.
AC O24133;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

20-AUG-2001 (Rel. 40, Last annotation update)
MAGNESIUM-CHELATASE SUBUNIT CHLD PRECURSOR (MG-PROTOPORPHYRIN IX
CHELATASE) (MG-CHELATASE SUBUNIT D).
CHLD.
GN Nicotiana tabacum (Common tobacco).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
NCBI_TaxID=4097;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=98079235; Pubmed=9418040;
RX Paperbrock J., Graefe S., Kruse E., Haenel F., Grimm B.;
RA "Mg-chelatase of tobacco: identification of a Chl D cDNA sequence
RT encoding a third subunit, analysis of the interaction of the three
RT subunits with the yeast two-hybrid system, and reconstitution of the
RT enzyme activity by co-expression of recombinant Chl D, Chl H and Chl
RT I.";
RL Plant J. 12:981-990(1997).
CC -1- FUNCTION: INVOLVED IN CHLOROPLAST PIGMENT BIOSYNTHESIS; INTRODUCES
CC A MAGNESIUM ION INTO PROTOPORPHYRIN IX TO YIELD MG-
CC PROTOPORPHYRIN IX.
CC -1- PATHWAY: CHLOROPHYLL BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.
CC -1- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNITS D/I FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Y10022; CAA71128.1; "-
DR InterPro: IPR000523; Mg_chelatase_chII.
DR InterPro: IPR002035; VWFA.
DR Pfam: PF01078; Mg_chelatase; 1.
DR SMART: SM00327; VWFA; 1.
DR PROSITE: PSS0234; VWFA; 1.
KW Photosynthesis; Chlorophyll biosynthesis; Chloroplast;
KW Transit peptide.
FT TRANSIT 1 62 CHLOROPLAST (POTENTIAL).
FT CHAIN 63 758 MAGNESIUM-CHELATASE SUBUNIT CHLD.
FT DOMAIN 556 752 VWFA.
FT DOMAIN 399 445 GLU/PRO-RICH.
FT DOMAIN 409 416 POLY-PRO.
FT DOMAIN 427 434 POLY-GLU.
SQ SEQUENCE 758 AA; 83012 MW; ED531E85D5FAFEF8 CRC64;

Query Match 6.2%; Score 87; DB 1; Length 758;
Best Local Similarity 21.6%; Pred. No. 13;
Matches 50; Conservative 3%; Mismatches 93; Indels 54; Gaps 9;

	Query Match	6.28;	Score 87;	DB 1;	Length 758;
	Best Local Similarity	21.68;	Pred. No. 13;	Mismatches	93; Indels 54; Gaps
	Matches	50;	Conservative	35;	
Qy	24	SDEPTKLSPATGATLAFVDFVTGSMDDLMQVIDGAS-RILERSLSSRSRVIANXALV	82		
Db	542	TDMRAKMRKAG-ALVFVVDSGSMAINRMQNAKGAAKLAEISTSDQV----	CII 596		
Qy	83	PFHPDPIGPVTLTADPVV--FQREQLLYVOGGDCPEMSVGAKAAVEVANPGS-----	F 136		
Db	597	PFPG-DAAEVLLPPSRISIMARNLERPCGGGSPLAHGLTTAVRVGNMNAEKSGDVGRIM	655		
Qy	137	IYVFSDARA-----KDYHKKNELLQL-----	LQLKQSQVV 166		
Db	656	IVALTDRANISLRSTDPAEAESDAPRPSQEELKDEILEVAGKIYKTGMSELLVIDTENK	715		
Qy	167	FVLFGDCGDTRHPGYLAFEETIASTSSGOVFOLDKQQQVSEVLKKWVESATOASK	218		
Db	716	FVSTGFA-----KETARVAQKQKYLYLPNASDAVISAAAYKDALSAULK	756		

```

RESULT 14
STRP_STRQ  STANDARD;          PRT;   440 AA.
AC  P00779;
DT  21-JUL-1986 (Rel. 01, Created)
DT  01-APR-1988 (Rel. 07, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  STREPTOKINASE C PRECURSOR.
GN  SKC.
OS  Streptococcus equisimilis.
OC  Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC  Streptococcus.
OX  NCBI_TaxID=119602;
[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=H46A;
RX  MEDLINE=85232082; PubMed=2989113;
RA  Malke H., Roe B., Ferretti J.J.;
RT  "Nucleotide sequence of the streptokinase gene from Streptococcus
FT  equisimilis H46A.";
RL  Gene 34:357-362(1985).
[2]
RN  SEQUENCE OF 27-440.
RP  MEDLINE=83127125; PubMed=6760891;
RA  Jackson K.W., Tang J.;
RX  "Complete amino acid sequence of streptokinase and its homology with
RT  serine proteases.";
RL  Biochemistry 21:6620-6625(1982).
CC  -!- FUNCTION: THIS PROTEIN IS NOT A PROTEASE, BUT IT ACTIVATES
CC  PLASMINOGEN BY COMPLEXING WITH IT. AS POTENTIAL VIRULENCE FACTOR,
CC  IT IS THOUGHT TO PREVENT THE FORMATION OF EFFECTIVE FIBRIN
CC  BARRIERS AROUND THE SITE OF INFECTION, THEREBY CONTRIBUTING TO THE
CC  INVASIVENESS OF THE CELLS.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
EMBL; K02986; AAA26974.1; -.
EMBL; X72832; CAA51351.1; -.
PIR; A00967; B2SO.
PIR; A22801; A22801.
KW  Plasminogen activation; Signal; Virulence.
FT  SIGNAL
FT  CHAIN          1 26          STREPTOKINASE C.
FT  VARIANT        27 440          L -> D.
FT  VARIANT        195 195          L -> D.
FT  VARIANT        207 207          D -> L.
FT  CONFLICT        298 300          EKY -> LEYK (IN REF. 2).
FT  CONFLICT        438 438          N -> D (IN REF. 2).
SQ  SEQUENCE      440 AA;  50140 MW;  8FC1F22648ACC77A CRC64;

Query Match          6.2%;   Score 86.5;  DB 1;   Length 440;
Best Local Similarity 24.9%;   Pred. No. 6.9;
Matches 56;  Conservative 31;  Mismatches 87;  Indels 51;  Gaps 10;

QY  52  DDLMOVDGASRIELSSRSRVIANVLPFHDPDTPVTLTADPV-----VFQR 103
DB  121  DLYFEVIDFAS---DATIDRN-----GKVFADKD-GSVTLRTPQVQEFLLSGHVRV 170
QY  104  ELRLQYVGGGDCPMSVGAIAKAAVEVANPGSVIVFSDAKADYHKNNELLQLQLKQ- 162
DB  171  PYKEPIQNA-----KSDVETVYQVTFPLNPDFFRPGDKTKLLKTLAIGDT 219
QY  163  ---SQVFEVLGTGCDRTHPGYLAPEIAS--TSSQVFOLDKQVSEVLKWWESAQASK 218
DB  220  ITSQELLAQAQSILKNHPGTYIYERDSIVTHNDIFRTILPMDQEFYRVKNREQAYR 279
QY  219  VHLLSADHEE-----EGEHTWRIPDPDS-LKEVTI 247

```

DB 280 INKKSGLNEINNTDLISEKYVVLKGEKPYD-PFDRSHLKFLTI 323
RESULT 15
PTEN HUMAN
ID PTEN HUMAN STANDARD; PRT; 403 AA.
AC O00633; O02679;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE PTEN (EC 3.1.3.48) (MUTATED IN MULTIPLE
DE ADVANCED CANCERS 1)
GN PTEN OR MMAC1 OR TEP1.
OS Homo sapiens (Human), and Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606, 9615;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS.
RC SPECIES=Human, and C.familiaris;
RX MEDLINE=97245711; PubMed=9090379;
RA Steck P.A., Pershouse M.A., Jasser S.A., Lin H., Yung W.K.A.,
RA Ligon A.H., Langford L.A., Baumgard M.L., Hattier T., Davis T.,
RA Frye C., Hu R., Swedlund B., Teng D.H.F., Tavtigian S.V.;
RT "Identification of a candidate tumour suppressor gene, MMAC1, at
RT chromosome 10q23.3 that is mutated in multiple advanced cancers.";
RL Nat. Genet. 15:356-363(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RX MEDLINE=97228181; PubMed=9072974;
RA Li J., Yen C., Liaw D., Podsypanina K., Bose S., Wang S.I., Puc J.,
RA Milaresis C., Rodgers L., McCombie R., Bigner S.H., Giovanello B.C.,
RA Ittmann M., Tycko B., Hibshoosh H., Wigler M.H., Parsons R.;
RT "PTEN, a putative protein tyrosine phosphatase gene mutated in human
RT brain, breast, and prostate cancer.";
RL Science 275:1943-1947(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RX MEDLINE=97330649; PubMed=9187108;
RA Li D.M., Sun H.;
RT "TEP1, encoded by a candidate tumor suppressor locus, is a novel
RT protein tyrosine phosphatase regulated by transforming growth factor
RT beta.";
RL Cancer Res. 57:2124-2129(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RA Wang S., Li J., Liaw D., Bose S., Podsypanina K., Parsons R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97404346; PubMed=9256433;
RA Myers M.P., Stolarov J.P., Eng C., Li J., Wang S.I., Wigler M.H.,
RA Parsons R., Tonks N.K.;
RT "p-PTEN, the tumor suppressor from human chromosome 10q23, is a dual-
RT specificity phosphatase.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9052-9057(1997).
RN [6]
RP VARIANT CD ASN-137 INS.
RX MEDLINE=98007985; PubMed=9345101;
RA Tsou H.C., Teng D.H.-F., Ping X.L., Brancolini V., Davis T., Hu R.,
RA Xie X.X., Gruener A.C., Schrager C.A., Christiano A.M., Eng C.,
RA Steck P., Ott J., Tavtigian S.V., Peacocke M.;
RT "The role of MMAC1 mutations in early-onset breast cancer: causative
RT in association with Cowden syndrome and excluded in BRCA1-negative
RT cases.";
RL Am. J. Hum. Genet. 61:1036-1043(1997).
RN [7]
RP VARIANTS CD GLU-343 AND LEU-347.
RX MEDLINE=98153621; PubMed=9399897;

RA Lynch E.D., Ostermeyer E.A., Lee M.K., Arena J.F., Ji H., Dann J.,
RA Swishhelm K., Suchard D., Macleod P.M., Kvinnslund S., Gjertsen B.T.,
RA Helmdal K., Lubs H., Moeller P., King M.-C.;
RT "Inherited mutations in PTEN that are associated with breast cancer,
RT Cowden disease, and juvenile polyposis.";
RL Am. J. Hum. Genet. 61:1254-1260(1997).
RN [8]
RP VARIANTS CD ARG-123 AND ARG-124.
RX MEDLINE=97402224; PubMed=9259288;
RA Nelen M.R., van Staveren W.C.G., Peeters E.A.J., Ben Hassel M.,
RA Gorlin R.J., Hamm H., Lindboe C.F., Fryns J.-P., Sijmons R.H.,
RA Woods D.G., Mariman E.C.M., Padberg G.W., Kremer H.;
RT "Germline mutations in the PTEN/MMAC1 gene in patients with Cowden
RT disease.";
RL Hum. Mol. Genet. 6:1383-1387(1997).
RN [9]
RP VARIANT CD GLU-129.
RX MEDLINE=97285123; PubMed=9140396;
RA Liaw D., Marsh D.J., Li J., Dahia P.L.M., Wang S.I., Zheng Z.,
RA Bose S., Call K.M., Tsou H.C., Peacocke M., Eng C., Parsons R.;
RT "Germline mutations of the PTEN gene in Cowden disease, an inherited
RT breast and thyroid cancer syndrome.";
RL Nat. Genet. 16:64-67(1997).
RN [10]
RP VARIANT BZS ARG-170.
RX MEDLINE=97385233; PubMed=9241266;
RA Marsh D.J., Dahia P.L.M., Zheng Z., Liaw D., Parsons R., Gorlin R.J.,
RA Eng C.;
RT "Germline mutations in PTEN are present in Bannayan-Zonana syndrome.";
RL Nat. Genet. 16:333-334(1997).
RN [11]
RP VARIANT CD GLU-289.
RX MEDLINE=99014194; PubMed=9797362;
RA Chi S.-G., Kim H.-J., Park B.-J., Min H.-J., Park J.-H., Kim Y.-W.,
RA Dong S.-H., Kim B.-H., Lee J.-I., Chang Y.-W., Chang R., Kim W.-K.,
RA Yang M.-H.;
RT "Mutation abrogation of the PTEN/MMAC1 gene in gastrointestinal
RT polyps in patients with Cowden disease.";
RL Gastroenterology 115:1084-1089(1998).
RN [12]
RP VARIANTS CD HIS-68 AND PRO-112.
RX MEDLINE=98260873; PubMed=9600246;
RA Tsou H.C., Ping X.L., Xie X.X., Gruener A.C., Zhang H., Nini R.,
RA Swisshelm K., Sybert V., Diamond T.M., Sutphen R., Peacocke M.;
RT "The genetic basis of Cowden's syndrome: three novel mutations in
RT PTEN/MMAC1/TEP1.";
RL Hum. Genet. 102:467-473(1998).
RN [13]
RP VARIANTS CD AND BZS.
RX MEDLINE=98133933; PubMed=9467011;
RA Marsh D.J., Coulon V., Lunetta K.L., Rocca-Serra P., Dahia P.L.M.,
RA Zheng Z., Liaw D., Caron S., Duboue B., Lin A.Y., Richardson A.,
RA Bonnetblanc J.-M., Bressieux J.-M., Cabarro-Moreau A., Chompert A.,
RA Denange L., Eeles R.A., Yehanda A.M., Fearon E.R., Fricker J.-P.,
RA Gorlin R.J., Hudson S.V., Lacombe D., Leprat F., Odent S.,
RA Toulouse C., Olopade O.I., Sobol H., Tishler S., Woods C.G.,
RA Robinson B.G., Weber H.C., Parsons R., Peacocke M., Longy M., Eng C.;
RT "Mutation spectrum and genotype-phenotype analyses in Cowden disease
RT and Bannayan-Zonana syndrome, two hamartoma syndromes with germline
RT PTEN mutation.";
RL Hum. Mol. Genet. 7:507-515(1998).
RN [14]
RP VARIANT CD TYR-136.
RX MEDLINE=98408002; PubMed=9735393;
RA Scala S., Bruni P., Lo Muzio L., Mignogna M., Viglietto G., Fusco A.;
RT "Novel mutation of the PTEN gene in an Italian Cowden's disease
RT kindred.";
RL Int. J. Oncol. 13:665-668(1998).
RN [15]
RP VARIANT JPC ARG-35.
RX MEDLINE=98085856; PubMed=9425889;
RA Olschwang S., Serova-Sinilnikova O.M., Lenoir G.M., Thomas G.;
RT "PTEN germ-line mutations in juvenile polyposis coli.";

```

QY 172 --DCGDRTHRGYLAFEBIASSTSSGVQFOLDKQVSEVLKWWESAIOASKVHLLSADHEEE 2239
      : : : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 283 PEETSEKVENGLCDQBEIDSICSTERADNDKEYL-----VLTLTKNLLDKANKDKA 333
      : : : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 230 GEHTWRIPFPDPSLK 243
      : | | : | |
Db 334 NRY-----FSPNEK 342

```

```

Query Match      6.1%; Score 86; DB 1; Length 403;
Best Local Similarity 21.1%; Pred. No. 6.7;
Matches 41; Conservative 41; Mismatches 76; Indels 36; Gaps 8;

Qy 68 SLSSRSRIANYALVPHPDPIGPVTLTADPVFQRELRLQLYVOGGDC-PEMSVGAIIKA 126
   : : | : : : | | | : : : | | : | : | : | : | : | : | : | : | : |
Db 167 TIPSQRRVYVYVYLLKNHLDYRPVALLFHMMFEF---TIPMFSGGTCNFQVVCQLKV 222

Qy 127 AVEVANPG-----SFYI-----VPSDARANDYHKKNELLQLLQLKQSQV-VFVLITG 171
   : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 223 KIYSSNGPTRRDKFMVFEEPPQLPVCGDITKVFHFQKNRLMKDKRHFHWVNTFFIPG 282

```

...

GenCore version 4.5.
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 21:15:02 ; Search time 65.89 Seconds
(without alignments)
610.486 Million cell updates/sec

Title: US-09-665-728-1
Perfect score: 1403
Sequence: 1 MTPRAQLPLLATYVWAA.....IEVRDPLGMSQSPPLLMQD 275

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL17.*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhmc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	628.5	44.8	5198	5 076518	076518 caenorhabdi
2	282.5	20.1	536	4 Q9UMP9	Q9UMP9 homo sapien
3	282.5	20.1	852	4 Q9Y334	Q9Y334 homo sapien
4	282.5	20.1	873	11 Q9Z108	Q9Z108 mus musculu
5	282.5	20.1	891	11 Q9JHA8	Q9JHA8 mus musculu
6	116	8.3	353	11 Q9QXEL	Q9QXEL mus musculu
7	110	7.8	2229	5 Q19853	Q19853 caenorhabdi
8	109	7.8	2014	5 Q22774	Q22774 caenorhabdi
9	106	7.6	2470	13 Q90681	Q90681 gallus gall
10	105.5	7.5	1363	1 Q9HNT5	Q9HNT5 halobacteri
11	105	7.5	476	5 Q23410	Q23410 caenorhabdi
12	105	7.5	2214	5 Q20219	Q20219 caenorhabdi
13	101.5	7.2	1324	2 Q52048	Q52048 streptomyc
14	96.5	6.9	1292	10 Q9M109	Q9M109 arabidopsis
15	93.5	6.7	1576	2 Q05201	Q05201 rhodococcus
16	93.5	6.7	1576	2 Q69354	Q69354 rhodococcus
17	93	6.6	458	5 Q93377	Q93377 caenorhabdi
18	93	6.6	861	11 Q9EQK5	Q9EQK5 mus musculu
19	93	6.6	1362	1 Q9V294	Q9V294 pyrococcus

20	92.5	6.6	510	8 Q9G8W0	Q9G8W0 rhodomonas
21	92	6.6	584	2 Q9XCV7	Q9XCV7 streptomyc
22	92	6.6	611	5 Q9NLD1	Q9NLD1 caenorhabdi
23	92	6.6	788	13 Q07012	Q07012 xenopus lae
24	92	6.6	1648	11 Q9Q205	Q9Q205 mus musculu
25	92	6.6	1944	5 Q9N375	Q9N375 caenorhabdi
26	91.5	6.5	414	2 Q53284	Q53284 streptococc
27	91	6.5	1250	5 Q23409	Q23409 caenorhabdi
28	90.5	6.5	392	10 Q40742	Q40742 oryza sativ
29	89.5	6.4	772	13 Q9PU04	Q9PU04 ictalurus p
30	89	6.3	537	2 Q9AB06	Q9AB06 caulobacter
31	89	6.3	1716	5 Q9VZG5	Q9VZG5 drosophila
32	88.5	6.3	547	2 Q9KR88	Q9KR88 vibrio chol
33	88.5	6.3	869	10 Q9FW67	Q9FW67 oryza sativ
34	88	6.3	403	4 Q43460	Q43460 homo sapien
35	88	6.3	789	5 Q9BKQ4	Q9BKQ4 caenorhabdi
36	88	6.3	1229	10 Q49749	Q49749 arabidopsis
37	88	6.3	1229	10 Q9SY12	Q9SY12 arabidopsis
38	88	6.3	1229	10 Q9FWX8	Q9FWX8 arabidopsis
39	88	6.3	1278	10 Q9FWX7	Q9FWX7 arabidopsis
40	87.5	6.2	552	1 Q28762	Q28762 archaeoglob
41	87	6.2	353	2 P72409	P72409 streptomyc
42	87	6.2	417	11 Q9Z0R1	Q9Z0R1 cavia porce
43	87	6.2	791	6 Q9GK49	Q9GK49 bos taurus
44	87	6.2	3722	2 P94873	P94873 lysobacter
45	86.5	6.2	1230	2 Q59325	Q59325 clostridium

ALIGNMENTS

RESULT 1
076518 PRELIMINARY; PRT: 5198 AA.
ID O76518 Q10036;
AC O76518: Q10036;
DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE HEMICENTIN PRECURSOR.
GN HIM-4 OR F15G9.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Sulston J., Kershaw J.;
RT "Hemicentin is Required for Hemidesmosome Mediated Cell Adhesion and Germ-Line Chromosome Segregation in C. elegans.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Sulston J., Kershaw J.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS; ISOFORM F15G9.4A AND ISOFORM F15G9.4B (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: TO IMMUGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX DOMAIN.
DR EMBL; AF074901; AAC26792.1; -;
DR EMBL; Z47068; CAA87336.1; -;
DR EMBL; Z47070; CAA87336.1; JOINED.
DR EMBL; Z47070; CAA87345.1; -;
DR EMBL; Z47068; CAA87345.1; JOINED.
DR EMBL; Z47068; CAA87335.1; -;
DR EMBL; Z47070; CAA87335.1; JOINED.
DR EMBL; Z47070; CAA87344.1; -;
DR EMBL; Z47068; CAA87344.1; JOINED.
DR HSSP; P35555; IEMN.
DR WormPep; F15G9.4A; CE18595.
DR WormPep; F15G9.4B; CE18596.
DR InterPro; IPR000152; Asx_hydroxyl.

DR InterPro: IPR000515; BPD_transp.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-like.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig-like.
 DR InterPro: IPR003006; Ig_LHC.
 DR InterPro: IPR002035; WFA.
 DR Pfam: PF00047; Ig_47.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00408; IGC2; 44.
 DR SMART: SM00410; IGC-like; 3.
 DR SMART: SM00327; VWA; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE: PS00402; BPD_TRANS INN_MEMBER; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 DR PROSITE: PS01187; EGF_CA; 2.
 KW Alternative splicing; Calcium-binding; EGF-like domain; Glycoprotein;
 KW Repeat; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 5198 HEMICENTIN.
 FT VARSPLIC 1462 5198 MISSING (IN ISOFORM F15G9_4A).
 SQ SEQUENCE 5198 AA; 570816 MW; DA85111F2B58D37B CRC64;

Query Match 44.8%; Score 628.5; DB 5; Length 5198;
 Best Local Similarity 48.6%; Pred. No. 1.7e-43;
 Matches 126; Conservative 51; Mismatches 69; Indels 13; Gaps 3;

QY 10 LLLATYVVAATVDEPTKTLSPATGDTAFVFDVTCGSMWDDLMQVIGASRLERSL 69
 DB 15 LLLAT---TCSSVNDK-----NDPTGKSSLAFAVFDITGSMFDDLVQVREGAIFKTYM 66
 QY 70 SSRSRVIANVLPHPDPDIPGVTLTADPVFORLRLYVGGDGPMSVGAIAAIVE 129
 DB 67 AQRELIYNYINVPHDPVLGLINTDSTYFNRLSKYVYVGGDGPCKTITGLKALQ 126
 QY 130 VANPGSFYVDSARAKYHKKNELQLQLKQSVFVLTGDCGDRTHPGYLAFFETIAS 189
 DB 127 ISLPSFYVFTDARKYHLEDEVLNTIQEQSSVFMVMTDCGNRTHPGFRTVEKIAA 186
 QY 190 TSSGVQFOLDKQVSEVLKWSAIOAKVHLLSADHEEGHTWRIPDPSPKVTISL 249
 DB 187 ASFGQVHLEKSDVSTVLEYVRAVQKQVHLMYERARGGTVSINIPVDKHLSELTISL 246
 QY 250 SG-----FGPETEVRDPLG 263
 DB 247 SGDKDDSDNLDIVLRDPEG 265

RESULT 2
 ID Q9UMP9 PRELIMINARY; PRT; 536 AA.
 AC Q9UMP9;
 DT 01-MAY-2000 (TremBLrel. 13, Created)
 DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
 DE G7C PROTEIN (FRAGMENT).
 GN G7C.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Aguado B.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA Albertella M.R.; Campbell R.D.;
 RT "Characterisation of the novel gene G7c located in the class III
 region of the human Major Histocompatibility Complex";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96411681; PubMed=8812450;
 RA Albertella M.A.; Jones H.; Thomson W.; Olavesen M.G.; Campbell R.D.;
 RT "Localization of eight additional genes in the human major
 RT histocompatibility complex, including the gene encoding the casein
 RT kinase II beta subunit (CSNK2B).";
 RL Genomics 36:240-251(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Albertella M.R.;
 RL Thesis (1997); University of Oxford, Department of Biochemistry,
 RL MRC Immunochimistry Unit.
 DR EMBL: AJ245418; CAB52192.1;
 FT NON_TER 1
 FT NON_TER 536
 SQ SEQUENCE 536 AA; 57837 MW; B32D8DD8E24BD06D CRC64;

Query Match 20.1%; Score 282.5; DB 4; Length 536;
 Best Local Similarity 29.2%; Pred. No. 1.3e-15;
 Matches 79; Conservative 50; Mismatches 93; Indels 49; Gaps 7;

QY 25 DEPTKTLSPATGDTAFVFDVTCGSMWDDLMQVIGASRLERSLSRSR----VIANYA 80
 DB 3 DRDFSRLLDITPASSLSFLVDTTSGMGEI-----NAKIQARHLVQRRGSPMEPVHYV 57
 QY 81 LVPFHPDIPGVTLTADPVFORLRLYVGGDGPMSVGAIAAIVEANPGSTIYVF 140
 DB 58 LVPFHPDIPGVTLTADPVFORLRLYVGGDGPMSVGAIAAIVEANPGSTIYVF 117
 QY 141 SDARADYHKKNELQLQLKQSVFVLTGDCG---DRTHPGYLA-----FEETAS 191
 DB 118 TDASPKDAFLTNQVSLTQERRCRVFTVETDSVQGRARRRILSPLEFPYKAVLAS 177
 QY 192 SGVQFOLDKQVSEVLKWSAIOAKVHLLSADHEEGHTWRIPDPFS-----241
 DB 178 GGEVFTKQHIRDVAIVGESMAA---LVT-----LPLDPVVPVPGPLVF 221
 QY 242 -----LKEVTISLSGPGPEIEVRDPLGMSQG 267
 DB 222 SVDGILLQKTVIRIHGDISSFWINKPAGVSQG 252

RESULT 3
 ID Q9Y334 PRELIMINARY; PRT; 852 AA.
 AC Q9Y334;
 DT 01-NOV-1999 (TremBLrel. 12, Created)
 DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
 DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
 DE NG37.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rowen L.; Qin S.; Madan A.; Dickhoff R.; Dors M.; Madan A.; Hicks P.;
 RA Loretz C.; Ratcliffe A.; Abbasi N.; Shaffer T.; Hood L.;
 RT "Sequence of the human major histocompatibility complex class III
 region";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF134726; AAD21820.1;
 SQ SEQUENCE 852 AA; 92538 MW; 025430B3F912A941 CRC64;

Query Match 20.1%; Score 282.5; DB 4; Length 852;
 Best Local Similarity 29.2%; Pred. No. 2.5e-15;
 Matches 79; Conservative 50; Mismatches 93; Indels 49; Gaps 7;

QY 25 DEPTKTLSPATGDTAFVFDVTCGSMWDDLMQVIGASRLERSLSRSR----VIANYA 80

Db 301 DRFSRLDITPASSLSFVLTGSMGEEI-----NAAKIQARHLVEQRRGSPMEPVHYV 355

Qy 81 LVPFHDPIGPVTLTADPVFQRELRLQLYVOGGDCPEMSVGAIKAAVEANPGSFIVVF 140

Db 356 LVPFHDGPGPVFTTSDPDSFQWQNEIHALGGDEPEMCLSAQLALLHTPPLSDIFVF 415

Qy 141 SDARKYHKKNELLQLLQKOSQVVFVLTGCG---DRTHPGYLA-----FEEIASTS 191

Db 416 TDASPKDAFLTNVESLTQERRCRVTLVTDTSRQGRARREILSPURFEPYKAVALAS 475

Qy 192 SGOVQLDKQOVSEVLKWKVESAIQASKVHLLSADHEEGEHTWRIPFDPDS----- 241

Db 476 GGEVFTTKDQHIDRVAALVIGESMAA-----LVT-----LPLDPPVVVPQPLVF 519

Qy 242 -----LKEVTISLSGPGPEIEVRDPLGMSQ 267

Db 520 SVDGLQKITVRIHGDISSFWIKNPAGVSQ 550

RESULT 4

ID Q921Q8 PRELIMINARY; PRT; 873 AA.

AC Q921Q8;

DT 01-MAY-1999 (Tremblrel. 10, Created)

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE HYPOTHETICAL 94.9 KDA PROTEIN.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RC STRAIN=129;

RA Rowen L., Madan A., Qin S., Shaffer T., Ratcliffe A., Abbasi N.,

RA Dickhoff R., James R., Loretz C., Lasky S., Hood L.;

RT "Sequence of the mouse major histocompatibility locus class III

RT region.";

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF109905; AAC84152.1; -

DR InterPro: IPR002035; vWFA.

DR SMART; SM00327; vWA; 1.

KW Hypothetical protein.

SQ SEQUENCE 873 AA; 94894 MW; 5761E3620F8D831C CRC64;

Query Match 20.1%; Score 282.5; DB 11; Length 873;

Best Local Similarity 30.3%; Pred. No. 2.6e-15;

Matches 81; Conservative 45; Mismatches 100; Indels 41; Gaps 6;

Qy 25 DEPTKTLSPATGDATLAFVFDVTGSMDDLMQVIDGASRILERSLSRSRVIANVLPVF 84

Db 310 DKAFSRLDITPASSLSFVLTGSMGEEINAAKIQARRIVEQRRGSPMEPVF-YILVPF 368

Qy 85 HDPIGVPVTLTADPVFQRELRLQLYVOGGDCPEMSVGAIKAAVEANPGSFIVVFSDAR 144

Db 369 HDPGFGPVFTTSDPDSFQWQNEIHALGGDEPEMCLSALEALLHTPPLSDIFVFTDAS 428

Qy 145 AKDYHKKNELLQLLQKOSQVVFVLTGD---CGDRTHPGYLA-----FEEIASTSSGOV 195

Db 429 PKDALLNRVSLTRRCRVTFLVTDPSRTGRRRRREALSPURFEPYEAIRASGGEV 488

Qy 196 FQLDKQOVSEVLKWKVESAIQASKVHLLSADHEEGEHTWRIPFDPDS----- 241

Db 489 IFTRDQYIQD-----VAAVIGESMAGLVT-----LPLDPPVFTPGPCPVFSDS 532

Qy 242 -LKEVTISLSGPGPEIEVRDPLGMSQ 267

Db 533 LLWQVTRMHGDISFSFWIKSPAGVSQ 559

RESULT 5

Q9JHA8

ID Q921Q8 PRELIMINARY; PRT; 891 AA.

AC Q9JHA8;

DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE G7C (G7C PROTEIN).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RC STRAIN=C57BL/RIJ, AND BALB/C; TISSUE=BRAIN;

RC MEDLINE=98211706; PubMed=9551980;

RA Snoek M., Teuscher C., van Vugt H.;

RT "Molecular analysis of the major MHC recombinational hot spot located

RT within the G7c gene of the murine class III region that is involved in

RT disease susceptibility.";

RL J. Immunol. 160:266-272(1998).

RN [2]

RC STRAIN=C57BL/RIJ, AND BALB/C; TISSUE=BRAIN;

RC MEDLINE=20260998; PubMed=10803853;

RA Snoek M., Albertella M.R., van Kooij M., Wixon J., van Vugt H.,

RA de Groot K., Campbell R.D.;

RT "G7c, a novel gene in the mouse and human major histocompatibility

RT complex class III region, possibly controlling lung tumor

RT susceptibility.";

RL Immunogenetics 51:383-386(2000).

DR EMBL; AF134318; AAF69177.1; -

DR EMBL; AF134319; AAF61401.1; -

DR InterPro: IPR002035; vWFA.

DR SMART; SM00327; vWA; 1.

SQ SEQUENCE 891 AA; 95983 MW; 1EA26B512118720 CRC64;

Query Match 20.1%; Score 282.5; DB 11; Length 891;

Best Local Similarity 30.3%; Pred. No. 2.7e-15;

Matches 81; Conservative 45; Mismatches 100; Indels 41; Gaps 6;

Qy 25 DEPTKTLSPATGDATLAFVFDVTGSMDDLMQVIDGASRILERSLSRSRVIANVLPVF 84

Db 302 DKAFSRLDITPASSLSFVLTGSMGEEINAAKIQARRIVEQRRGSPMEPVF-YILVPF 360

Qy 85 HDPIGVPVTLTADPVFQRELRLQLYVOGGDCPEMSVGAIKAAVEANPGSFIVVFSDAR 144

Db 361 HDPGFGPVFTTSDPDSFQWQNEIHALGGDEPEMCLSALEALLHTPPLSDIFVFTDAS 420

Qy 145 AKDYHKKNELLQLLQKOSQVVFVLTGD---CGDRTHPGYLA-----FEEIASTSSGOV 195

Db 421 PKDALLNRVSLTRRCRVTFLVTDPSRTGRRRRREALSPURFEPYEAIRASGGEV 480

Qy 196 FQLDKQOVSEVLKWKVESAIQASKVHLLSADHEEGEHTWRIPFDPDS----- 241

Db 481 IFTRDQYIQD-----VAAVIGESMAGLVT-----LPLDPPVFTPGPCPVFSDS 524

Qy 242 -LKEVTISLSGPGPEIEVRDPLGMSQ 267

Db 525 LLWQVTRMHGDISFSFWIKSPAGVSQ 551

RESULT 6

Q9QXE1

ID Q9QXE1 PRELIMINARY; PRT; 353 AA.

AC Q9QXE1;

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)

DE SHD-1 PROTEIN.

GN SHD-1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;
RA Yoshida M., Kuwahara K., Sakaguchi N.;
RT "Molecular cloning of a novel gene, SHD-1, belonging to cell cycle-
RT related SAC3 family.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ131957; CAB65242.1;
SQ SEQUENCE 353 AA; 38203 MW; D486B7B301493447 CRC64;

Query Match      8.3%; Score 116; DB 11; Length 353;
Best Local Similarity 24.9%; Pred. No. 0.089;
Matches 70; Conservative 31; Mismatches 98; Indels 82; Gaps 13;

Qy 3 PRAQLL---PILLATYTVVAAVTSDEPTKTLSPATGDATLA---FVFDVTGSMWDDL- 54
Db 64 PPSLLRPPVLLATVRVILAGEVA-----GRGDVSCAEVASFVADRLRAVRLLDLS 113

Qy 55 MQVID--GASRILERSLSRSRVIANYPHDPDIPGVTL--TADPVVFQRE----- 104
Db 114 LOGVDADAATVLEAALATLLAVAR-----VRPEETRGAAADPVLLQTQVQEGFG 163

Qy 105 -LRQLYVOGGDCP-----EMSVGAIKAAVEVANPGSFYVFSARAK 146
Db 164 SLRRCYARGKGPYRQAFAOGLFLLYNLGSLAPAGGSTAALRACPLQAALAVDAARE 223

Qy 147 DYHKK-NELLQLQLKQSVVFLTGDCGDRTHPGYLAFEETA-----STSSGOVFOLD 199
Db 224 DNHARLRLRLTLPYLS-----CAVQEHGYARKALARLSRALSTPKGTLPID 274

Qy 200 KQVSEVLKVESATQASKVHLLSAD-----HEEG 230
Db 275 FTEHFLALDGLQEARLDLQAHGLTLDKRVVFLRQYSEEG 315

RESULT 7
Q19853 ID Q19853 PRELIMINARY; PRT; 2229 AA.
AC Q19853;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SIMILAR TO EGF-LIKE REPEAT. NCBI GI: 1125759.
GN F28B4.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden K., Sulston J.,
RA Thiermy-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Leimbach D.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

```
RA Waterston R.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U42834; AAA83584.1;
DR InterPro; IPR001442; C4;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001304; lectin_c.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00111; C4; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS0041; C_TYPELECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_6.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS02034; VWFA; 1.
DR EGF-like domain; Glycoprotein.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 2229 AA; 244874 MW; 873C82B6A97A25 CRC64;

Query Match      7.8%; Score 110; DB 5; Length 2229;
Best Local Similarity 20.1%; Pred. No. 4.4; 119; Indels 40; Gaps 9;
Matches 52; Conservative 48; Mismatches 119; Indels 40; Gaps 9;

Qy 39 TLAFVFDVTGSMWDDLMQVIDGASRILERSLSRSRVIANYPHDPDIPGVTL----- 94
Db 1551 TLLFVVEITRQSDTIVNQLIANKNIVTSA-----TNFA--PFWSYFGLVTFDITG 1600

Qy 95 -----TADPVVFQRELRLQLYVOGGDCPMSVGAIKAAVE-----VANPGSFYVFS 141
Db 1601 RTFEKYNVTTIDALITDLTAQSTAISTDGCASMPYLGVAHLLEHNDNVISIPNSEIFLVT 1660

Qy 142 DARAKDYHKKNELLQLKQSVVFLTGDCGDRTHPGYLAFE-----IASTSSGOVFOL 198
Db 1661 AAGPSDLNKYGEAMNSLENTEAHLHYIVYKSNCPTEFGVNVQDMTWLGCGSSGNILFT 1720

Qy 199 DKQVSEVL-KWVESATQASKVHLLS-----ADHE-ERGEHTWRIPDPSPKSVTISLSG 251
Db 1721 DSSNIVSLMNSYLPISLYGAS---ILQDPTGPANYSCTDGLPWFVPVDSNTFFIVVTSS 1777

Qy 252 PGPEIEVRDPLGMSOGSPP 270
Db 1778 EFGSLSVKDPDLQAHNVAP 1796

RESULT 8
Q22774 ID Q22774 PRELIMINARY; PRT; 2014 AA.
AC Q22774; O01703; O01984;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 221.3 KDA PROTEIN T25C12.3 IN CHROMOSOME X.
GN T25C12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Kershaw J.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: WEAK, TO C.ELEGANS ZK1193.2 AND TO THE C-TERMINAL OF
CC COLLAGENS.
DR EMBL; Z66566; CAA91487.1;
DR EMBL; Z78543; CAA91487.1; JOINED.
DR EMBL; Z93779; CAA91487.1; JOINED.
DR EMBL; Z93779; CAB07849.1;
DR EMBL; Z78543; CAB07849.1; JOINED.
DR EMBL; Z66566; CAB07849.1; JOINED.
```



```
QY 150 KKNELLQLLQKQSVVFLTGCGRTHPGYLAPEIASTSSGOVQFQDLKQOVSEVLKW 209
Db 451 -----TNATSATVHFVAADAGNFT-AGDLANASASVSKDIDMAPCKORVDVNLRW 501
QY 210 VESAIQASKVHLLSADHEEG 230
Db 502 VREVTDSNVTYVTTATDERG 522

RESULT 11
Q23410
ID Q23410 PRELIMINARY; PRT; 476 AA.
AC Q23410;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE SIMILARITY OVER A SHORT REGION TO TENASCIN PRECURSORS.
GN ZK1193.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Therry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Geisel C.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41553; AAA83292.1; -.
DR InterPro; IPR000561; EGF-like.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 476 AA; 53310 MW; F39C14163CC074DA CRC64;

Query Match 7.5%; Score 105; DB 5; Length 476;
Best Local Similarity 22.3%; Pred. No. 1.2;
Matches 53; Conservative 49; Mismatches 100; Indels 36; Gaps 10;

QY 38 ATAFVFDVTCMSMDLMOVIDGASRILE-----RSLSSRSRVIANVLPVPHD 86
Db 82 SAIAFVIRSSSKAQLNEAATNIVEYELHYFFQKFIILT---VVSNNAITSHSHE 138
QY 87 PDIGFVLTADPVVFORELRLQYV-OGGDCPEMSVGAIAKAAVEA----NPGSFIYVFS 141
Db 139 YDIGE-----DFANSIRSLVAPPTETECDDALLAGISKLTENSAFKYKPNPVPVFS 190
QY 142 DARAK-DVHKKNELLQLLQKQSVVFLT-----GDCG-DRTHPGYLAPEIASTSSGOV 195
Db 191 DGTANDFTTAGFMEQIVNTRAILFMITESAGSCNVDSVNTIFESRLSLQSLRGLL 250
QY 196 FQDKQOVSEVLKQVSEVFLTGCGRTHPGYLAPEIASTSSGOVQFQDLKQOVSEVLKW 251
```

```
Db 251 IOTSLMQLSDATFSAQDLWQYDT-ILTNLDECRKAPWFQPFVDSIDFLTLRASG 307

RESULT 12
Q20219
ID Q20219 PRELIMINARY; PRT; 2214 AA.
AC Q20219;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE CODED FOR BY C. ELEGANS CDNA YK131E8.3.
GN F40F4.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Therry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Wilson R.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U40420; AAA81430.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00032; vwa; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_6.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS02334; VVWA; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 2214 AA; 243435 MW; B0B387B67013A401 CRC64;

Query Match 7.5%; Score 105; DB 5; Length 2214;
Best Local Similarity 19.4%; Pred. No. 12;
Matches 51; Conservative 46; Mismatches 120; Indels 46; Gaps 8;

QY 39 TLAFFVDVTCMSMDLMOVIDGASRIILERSLSRSRVIANVLPVPHDPIGPVTL---- 94
Db 1548 SLLMVVETTRQNSDTVNQLLANKNIVSATVNNMPLMFTNGLVTF---DTTGRTFEKF 1604
QY 95 --TADPVVFORELRLQYV-OGGDCPEMSVGAIAKAAVEA----VANPGSFIYVFS 148
Db 1605 YTSIDDLITLTITQSNIAISTDGVCSMPYLGVLHLLHDDVDVIAIMPNSIEFLVTPAGPSDL 1664
QY 149 HKKNELLQLLQKQSVVFLT---TGDCG-----DRTHPGYLAPEIASTSSGOV 197
```

```
Db 1665 GNYVETMEVLENTQALHLYVYVSKTANCATFDGVNNVDMTWLGY-----GSSGNILF 1716
QY 198 LDKQVSEVLKWWESIAQSKV-----HLLSADHEEGEHTWRIFDPDSLKEVTS 248
Db 1717 TDPANIVNLFNFYLPYLGASVLQDPTGNTVTC-----DGLPWFVFPVDINTTPIYVT 1771
QY 249 LSGPGPEIEVRDPLGMSQSPPL 271
Db 1772 TSAEFGSLVKOPLGAHSAATPV 1794

RESULT 13
O52048 PRELIMINARY; PRT; 1324 AA.
AC O52048;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PEPTIDE SYNTHETASE MODULE (FRAGMENT).
GN CPS-1.
OS Streptomyces fradiae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1906;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A54145;
RA Hosted T.J.J., Baltz R.H.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016696; AAC01566.1; -.
DR HSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; DUF4.
DR InterPro; IPR003880; Phosphopant_attach.
DR Pfam; PF00501; AMP-binding; 1.
DR Pfam; PF00668; Condensation; 2.
DR Pfam; PF00550; pp-binding; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00075; ACP_DOMAIN; 1.
DR PROSITE; PS00455; AMP_BINDING; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 1324 1324
FT SEQUENCE 1324 AA; 140539 MW; 24285EDD3690964F CRC64;

Query Match 7.2%; Score 101.5; DB 2; Length 1324;
Best Local Similarity 20.6%; Pred. No. 11;
Matches 67; Conservative 38; Mismatches 123; Indels 97; Gaps 161;

QY 10 LLLATYVVAAYTSDEPTKTLSPAT-----GDATLAFVFDVTGSMWDDLMOV 57
Db 102 LLERALTAVSAPATPTAALAAATATARAAPRAAGCAPATIVDAFEARVATPEAPAV 161
QY 58 IDG-----ASRILERSLSRSRV---IANYALVPF-----HDP 87
Db 162 LAGGEELTYAELDARANRLARLLERGVGPESRVALTVRNWLPAVLGILKAGGCYVP 221
QY 88 -----DTGPVTLTADPVVFORELQRLVQGGDCPEMSVGAIKAAVEVAN 132
Db 222 VGATLPRERAARILRETAPVCLLTDPAEAAARTRT-APTGDDRDENAPGVE---RVVL 277
QY 133 PGSTIYVFSARAKDYHKKNELLOLLOLKQSOVVFLTGCGDRTHPGCYLAFEEIATSS 192
Db 278 TGALLAAPPAPPPDAERAGPLL-----PCHLAILLHTSGSS 314
QY 193 QGV--FQLDKQVSEVLKWWESIAQSKVHLLSADHEEGEHTWRIFDPDSLKEVTSLS 250
Db 315 GRPKGVAHEAQTALLSWAGTGVGADRLHRTVASTSES-----FDVSVFDTLVPLL 366
QY 251 GPGPEIEVRDPLGMSQ---GSPPLL 272
Db 367 TGGRIEIVENTLAVADRTGGEPSSL 391
```

```
RESULT 14
Q9MIQ9 PRELIMINARY; PRT; 1292 AA.
AC Q9MIQ9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE P-GLYCOPROTEIN-LIKE PROTEIN.
GN T17J13.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W.,
RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
CC TRANSPORTERS).
DR EMBL; ALI38651; CAB71875.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW ATP-binding; Transport.
SQ SEQUENCE 1292 AA; 139753 MW; 9C95DF3AC84E7432 CRC64;

Query Match 6.9%; Score 96.5; DB 10; Length 1292;
Best Local Similarity 29.2%; Pred. No. 27;
Matches 42; Conservative 15; Mismatches 50; Indels 37; Gaps 7;

QY 24 SDEPTKTLSPATGDATLAFVFP-----DVTGSMWDDLMOVIDGASRI--LERSLSRSR 74
Db 1034 SDESGRVLDMVKGDIELRHISFKYPSRPDV--QIFDLCLISIRAGKTIALVGGSGKST 1091
QY 75 VIANYALVPFHDPDIGPVTL-----TADPVVFORELQRLVQV-G 113
Db 1092 VIA--LLQRYFDPSGOITLDGVEIKTLQKLWLRQQTGLVSOEVLNETIRANIAVGK 1149
QY 114 GDCPEMSVGAIKAAVEVANPGSFI 137
Db 1150 GDATETE---IVSAAELSNAGFI 1170

RESULT 15
O05201 PRELIMINARY; PRT; 1576 AA.
AC O05201;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE RECEPTOR-LIKE HISTIDINE KINASE BPDS.
GN BPDS.
OS Rhodococcus sp.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1831;
RN [1]
```

Search completed: March 7, 2002, 22:07:47
Job time: 3165 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 22:05:20 ; Search time 48.12 seconds
(without alignments)
103.136 Million cell updates/sec

Title: US-09-665-728-1_COPY_56_122

Perfect score: 342

Sequence: 1 QVTDGASRLERSLSRSRV.....RELRLYVGSGDCPEMSVG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_1101.*

- 1: /SID22/gcgcdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SID22/gcgcdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SID22/gcgcdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SID22/gcgcdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SID22/gcgcdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SID22/gcgcdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SID22/gcgcdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SID22/gcgcdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SID22/gcgcdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SID22/gcgcdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SID22/gcgcdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SID22/gcgcdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SID22/gcgcdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SID22/gcgcdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SID22/gcgcdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SID22/gcgcdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SID22/gcgcdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SID22/gcgcdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SID22/gcgcdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SID22/gcgcdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SID22/gcgcdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SID22/gcgcdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	342	100.0	275	22 AAB83147	Rat secreted facto
2	175	51.2	3117	21 AAY53667	Sequence g1/32818
3	64	18.7	190	21 AAG52665	Arabidopsis thalia
4	64	18.7	229	21 AAG52664	Arabidopsis thalia
5	63.5	18.6	201	22 AAM41239	Human polypeptide
6	63	18.4	215	21 AAB19586	Human interleukin-
7	63	18.4	215	21 AAB19587	Mouse interleukin-
8	63	18.4	223	22 AAE00828	Human cardiotroph
9	63	18.4	225	19 AAW29715	Human neurotroph
10	63	18.4	225	19 AAW29716	Mouse neurotroph
11	63	18.4	225	19 AAW56141	Amino acid sequenc

Amino acid sequenc
Human cardiotroph
Human NNT-1 protei
Murine NNT-1 prote
Amino acid sequenc
Human protein sequ
Human lung tumour-
Human lung tumour-
Zea mays protein f
Arabidopsis thalia
Zea mays protein f
Zea mays protein f
Arabidopsis thalia
Arabidopsis thalia
Zea mays protein f
Mouse FCTR4 protei
A multifunctional
A multifunctional
Mouse growth facto
Mouse FCTR3 protei
Human VEGF-G prote
Human protein sequ
Zea mays protein f
Zea mays protein f
Zea mays protein f
Formate dehydrog
Breast and ovarian
Human polypeptide
Human polypeptide
Zea mays MAP kinas

ALIGNMENTS

RESULT 1	
AAB83147	ID AAB83147 standard; protein; 275 AA.
XX	AC AAB83147;
XX	DT 29-JUN-2001 (first entry)
XX	DE Rat secreted factor encoded by clone P00210D09.
XX	KW Rat; secreted factor; P00210D09; cardiac; nephrotropic;
KW	antiinflammatory; gene therapy; cardiac disease; renal disease;
KW	inflammatory disease.
OS	Rattus sp.
XX	XX Key Location/Qualifiers
FT	Peptide 1..21
FT	/label= Signal_peptide
FT	Protein 22..275
FT	/note= "Rat secreted factor"
FT	Domain 35..55
FT	/label= Transmembrane_domain
FT	Domain 123..143
FT	/label= Transmembrane_domain
XX	WO200123419-A2.
XX	PD 05-APR-2001.
XX	PF 27-SEP-2000; 2000WO-US26582.
XX	PR 27-SEP-1999; 99US-0156277.
XX	XX

```

PA (SCIO-) SCIOS INC.
XX Stanton LW, Kapoun AM;
XX WPI: 2001-328177/34.
XX DR N-PSDB: AAF82464.
XX
XX Novel secreted factor encoded by clone P00210D09 useful for diagnosing,
XX treating and/or preventing various cardiac, renal and inflammatory
XX diseases.
XX
XX Claim 9: Fig 1; 69pp; English.
XX
XX The present sequence is a novel secreted factor encoded by rat cDNA
XX clone P00210D09. The invention relates to a polypeptide comprising a
XX sequence of at least 80% identity to residues 22-122 of the present
XX sequence, or a sequence encoded by a nucleic acid hybridising under
XX stringent conditions to the complement of the coding region comprising
XX 1031 nucleotides, and having at least one biological activity of the
XX polypeptide encoded by clone P00210D09. The polypeptides and
XX polynucleotides of the invention are useful for the treatment of
XX cardiac, renal and inflammatory diseases. The polynucleotides are
XX useful in antisense mediated gene inhibition and in gene therapy.
XX The polypeptides are useful in assays for identifying lead compounds
XX that may be used as therapeutic agents in the treatment of cardiac,
XX kidney or inflammatory diseases.
XX
XX Sequence 275 AA;

Query Match 100.0%; Score 342; DB 22; Length 275;
Best Local Similarity 100.0%; Pred. No. 1.5e-38;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QVIDGASRLERSLSRSRVIANVALVPFHPDIPGVTLTADPVVFORLRLQYVQGGD 60
DB 56 qvidgasrlerslsrsrvianvalvpfhdipgvtltadpvvfqreirqlvygqgdd 115

QY 61 CPMSV 67
DB 116 cpmsv 122

RESULT 2
AY53667
ID AAY53667 standard; Protein: 3117 AA.
XX AC AAY53667;
XX DT 22-FEB-2000 (first entry)
XX DE Sequence gi/3328186 from an alignment with protein 608.
XX KW Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
XX KW bone development; gi/3328186.
XX OS Unidentified.
XX PN WO9960164-A1.
XX PD 25-NOV-1999.
XX PF 14-MAY-1999; 99WO-US11066.
XX PR 15-MAY-1998; 98US-0085673.
XX PA (QUARK-) QUARK BIOTECH INC.
XX PI Einat P, Mor O, Skalliter R, Feinstein E, Faerman A;
XX WPI: 2000-053304/04.
XX
XX Identification of stress induced genes for determining risk and

preventing, treating or controlling osteoporosis -
Claim 32: Fig 6A-R; 308pp; English.

The present sequence is obtained from a clustral X alignment with
protein 608. Protein 608 was identified using the method of the invention
after subjecting rat osteoblasts to mechanical stress. Expression of the
608 gene was found to be upregulated by about 3-fold in cells subjected
to mechanical strain. The specification describes a method for the
identification of genes responsive to a specific mechanical stress. The
method comprises applying the mechanical stress to an organism (tissue
or cells comprising bone cells), isolating the specific cellular
fractions and extracting mRNA from them, and differentially analysing the
mRNA in comparison with control samples. The method is used to identify
genes whose expression is responsive to a specific stress. The identified
genes are employed in determining risk associated with a physiological or
disease state. The risk determination methods are used for testing a
medicament for gene therapy. These medicaments, or genes identified by
the method of the invention, are used for treating, preventing or
controlling a physiological or disease state (especially osteoporosis or
bone density or other factors causing or contributing to osteoporosis or
its symptoms or other conditions involved in mechanical stress or its
lack. The methods can also be used for advancing research or studies in
bone development.

XX
XX Sequence 3117 AA;

Query Match 51.2%; Score 175; DB 21; Length 3117;
Best Local Similarity 45.5%; Pred. No. 2.2e-14;
Matches 30; Conservative 16; Mismatches 20; Indels 0; Gaps 0;

QY 1 QVIDGASRLERSLSRSRVIANVALVPFHPDIPGVTLTADPVVFORLRLQYVQGGD 60
DB 53 qvregaaikftvmaqrkellynlymvpfhdipgylgeinttdstvfmrqlskvyvhggdd 112

QY 61 CPMSV 66
DB 113 cpektl 118

RESULT 3
AGS2665
ID AAG52665 standard; Protein: 190 AA.
XX AC AAG52665;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 66973.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.

```

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

1

```
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 18.7%; Score 64; DB 21; Length 229;
Best Local Similarity 42.9%; Pred. No. 1.3;
Matches 21; Conservative 5; Mismatches 17; Indels 6; Gaps 2;

QY 6 ASRLERSLSRSRVIANVYALVPHDPDIPGVTLTADPVVFORELROLY 54
   ||: || ||| || : | || ||| | : | || |
Db 54 aprlnerilsslr--rsvaahpwhdleigp----gapqifnvvdriily 96

RESULT 5
AAM41239
ID AAM41239 standard; Protein; 201 AA.
XX AC AAM41239;
XX AC AAM41239;
DT 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 6170.
XX Human;
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX OS Homo sapiens.
XX PN W0200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-052317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang JT, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI60395.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
```

PS Example 2; SEQ ID NO 6170; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 201 AA;

Query Match 18.6%; Score 63.5; DB 22; Length 201;
Best Local Similarity 27.5%; Pred. No. 1.3;
Matches 19; Conservative 16; Mismatches 21; Indels 13; Gaps 3;

QY 10 LERSLSRRVIANVALVPF-----HDPDTPVLTADPVVQRELRLQLYVGGGD-- 60

DB 124 lmmvntgylaskacipylkkskvahlpnlsp-plnlpwwfkghcgrwvxvgxgdgl 182

QY 61 ---CPMSV 66

DB 183 cllicfeinl 191

RESULT 6

AAAB19586

ID AAB19586 standard; Protein; 215 AA.

XX AC AAB19586;

XX DT 22-JAN-2001 (first entry)

XX DE Human interleukin-B60 (IL-B60).

XX KW Interleukin-B60; IL-B60; human; cytokine; cytokine-like factor-1;
KW haematopoietic; inflammation; antiinflammatory; autoimmune disease;
KW therapy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..17

FT Protein /label= Signal_peptide

FT 18..215

FT /label= Mature-protein

XX WO200053631-A1.

XX 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-US06182.

XX PR 11-MAR-1999; 99US-0267901.

XX PA (SCHE) SCHERING CORP.

XX PI Oppmann B, Timans JC, Kastelein RA, Bazan JF;

XX DR WPI: 2000-587426/55.

XX DR N-PSDB; AAA88546.

XX PT Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes,

PT polypeptides, and nucleic acids, useful in research, diagnosis and for
PT treating inflammatory and autoimmune disorders -

XX Claim 1; Page 15-16; 97pp; English.

XX The present sequence is that of human interleukin-B60 (IL-B60), a
CC novel, small soluble cytokine-like protein that exhibits structural
CC motifs characteristic of a member of the long-chain cytokines, and
CC which shows homology to granulocyte colony stimulating factor and
CC interleukin-6. IL-60B may have either stimulatory or inhibitory
CC effects on haematopoietic cells, including e.g. lymphoid cells,
CC such as T-cells, B-cells, natural killer cells, macrophages,
CC dendritic cells, haematopoietic progenitors, etc. Methods are
CC provided for modulating the physiology or development of a cell or
CC tissue culture cells by contacting the cell with an agonist or
CC antagonist of IL-B60 or an agonist of antagonist of a complex of
CC mature IL-B60 and its partner, cytokine-like factor-1 (CLF-1, see
CC AAB19588). The IL-B60/CLF-1 cytokine serves as a key physiological
CC factor in motor neuron development and regeneration. IL-60B, its
CC agonists and antagonists may be used to treat inflammatory or
CC autoimmune disorders and also for drug screening.

XX Sequence 215 AA;

Query Match 18.4%; Score 63; DB 21; Length 215;
Best Local Similarity 38.9%; Pred. No. 1.6;
Matches 14; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 7 SRILERSLSRRVIANVALVPFHDPDTPVLTAD 42

DB 37 trylehqlrsilagtylnylgppfnepdfnprrlgae 72

RESULT 7

AAAB19587

ID AAB19587 standard; Protein; 215 AA.

XX AC AAB19587;

XX DT 22-JAN-2001 (first entry)

XX DE Mouse interleukin-B60 (IL-B60).

XX KW Interleukin-B60; IL-B60; mouse; cytokine; cytokine-like factor-1;
KW haematopoietic; inflammation; antiinflammatory; autoimmune disease;
KW therapy.

XX OS Mus musculus.

XX FH Key Location/Qualifiers

FT Peptide 1..17

FT Protein /label= Signal_peptide

FT 18..215

FT /label= Mature-protein

XX WO200053631-A1.

XX 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-US06182.

XX PR 11-MAR-1999; 99US-0267901.

XX PA (SCHE) SCHERING CORP.

XX PI Oppmann B, Timans JC, Kastelein RA, Bazan JF;

XX DR WPI: 2000-587426/55.

XX DR N-PSDB; AAA88547.

XX PT Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes,
PT polypeptides, and nucleic acids, useful in research, diagnosis and for

DR N-PSDB; AAD04201.
XX

XX
PS
Claim 12: Fig 3: 120pp: English.

Claim 12; Fig 3; 120pp; English.

(CLC) protein which is a member of the interleukin 6 (IL-6) cytokine family. The present sequence represents the human CLC protein. The present invention also describes screening methods for identifying agonists and antagonists of CLC activity, as well as methods for detecting cardiac and immune system-related disorders and therapeutic methods for treating cardiac and immune system-related disorders, e.g. heart failure, hypertension, cancers, autoimmune disorders and infections.

AA	Sequence	225 AA;
SQ		

Query Match 18.4%; Score 63; DB 20; Length 225;
Best Local Similarity 38.9%; Pred. No. 1.7;
Matches 14; Conservative 5; Mismatches 17; Indels

QY	7	SRILERSRSRVANYALVPFHDPDGPVLTAD	42
		: :	
Db	47	tryleqlrslagtylnylgppfnepdfnpprlgae	82

RESULT 14

AA87813
ID AA87813 standard; Protein; 225 AA.

AC AAY87813;

DT 24-AUG-2000 (first entry)

Human NNT-1 protein.

NNT-1; human; neurotrophic factor; nontropic; neuroprotective; treatment; anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological; nervous system degeneration; Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; Huntington's disease; peripheral neuropathy; neural retina degeneration; retinopathy; immune disorder; hematopoietic disorder.

OS Homo sapiens.

PN US6054294-A.

25-APR-2000.

12-DEC-1997; 97US-0988819.

03-FEB-1997; 97US-0792019

(AMGE-) AMGEN INC.

Chang M:

WPT: 2000-338492/29

OR N-PSDB; AAA39481.

New nucleic acids encoding neurotrophic factors useful for stimulating growth of motor or sympathetic neurons for treating neuron cell damage

XX
PS
Claim 1c; Fig 3; 42pp; English.

This invention describes a novel nucleic acid molecule (I) encoding a novel neurotrophic factor (NNT-1) (II) which has neurotropic, neuroprotective, anticonvulsant, antiparkinsonian, antidabetic and ophthalmological activity. (I) is useful for producing NNT-1 polypeptides which are useful for treating patients in whom various cells of the central, autonomic, or peripheral nervous system have degenerated and/or have been damaged by congenital disease, trauma, mechanical damage, surgery, stroke, ischemia, infection, metabolic disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1 proteins are used to treat diseases like Alzheimer's, Parkinson's, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's disease, peripheral neuropathy induced by diabetes or other metabolic

disorders, and/or dystrophies or degeneration of the neural retina such as retinitis pigmentosa, drug-induced retinopathies, stationary forms of night blindness, progressive cone-rod degeneration, immune disorders and hematopoietic disorders. (I) is effective in treating neurological conditions and promotes neuron regeneration. Neural functions are effectively restored in patients suffering from various neurological disorders. This sequence represents the human NNT-1 protein described in the method of the invention.

SQ Sequence 225 AA;

Query Match 18.4%; Score 63; DB 21; Length 225;
Best Local Similarity 38.9%; Pred. No. 1.7;
Matches 14; Conservative 5; Mismatches 17; Indels

Qy 7 SRILERSSSRRVIANVALVPFHDPIGPVLTAD 42
 : | | | | | : | | | | | :
Dd 47 trylehqlrslagtylnylgppfnepdfnprrlgae 82

RESULT 15

AAy87814
ID AAY87814 standard; Protein; 225 AA.

AC AAY87814;

24-AUG-2000 (first entry)

XX
DE
Murine NNT-1 protein.

NNT-1; neurotrophic factor; neurotropic; neuroprotective; treatment; anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological; KW nervous system degeneration; Alzheimer's disease; Parkinson's disease; KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; murine; KW Huntington's disease; peripheral neuropathy; neural retina degeneration; retinopathy; immune disorder; hematopoietic disorder.

Mus sp.

PN US6054294-A.

25-APR-2000.

12-DEC-1997; 97US-0988819.

03-FEB-1997: 97US-0792019

XX
PA
(AMGE-) AMGEN INC.

XX
pI Chang M:

XX
OR
WPI; 2000-338492/29.

OR N-PSDB; AAA39483.

New nucleic acids encoding neurotrophic factors useful for stimulating growth of motor or sympathetic neurons for treating neuron cell damage

IX
S
Claim 2b; Fig 5; 42pp; English.

This invention describes a novel nucleic acid molecule (I) encoding a neurotrophic factor (NNT-1) (II) which has neurotrophic, neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and ophthalmological activity. (I) is useful for producing NNT-1 polypeptides which are useful for treating patients in whom various cells of the central, autonomic, or peripheral nervous system have degenerated and/or have been damaged by congenital disease, trauma, mechanical damage, surgery, stroke, ischemia, infection, metabolic disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1 proteins are used to treat diseases like Alzheimer's, Parkinson's, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's disease, peripheral neuropathy induced by diabetes or other metabolic

CC disorders, and/or dystrophies or degeneration of the neural retina such
 CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of
 CC night blindness, progressive cone-rod degeneration, immune disorders and
 CC hematopoietic disorders. (I) is effective in treating neurological
 CC conditions and promotes neuron regeneration. Neural functions are
 CC effectively restored in patients suffering from various neurological
 CC disorders. This sequence represents the murine NNT-1 protein described in
 CC the method of the invention.
 XX
 SQ

Sequence 225 AA;

Query Match 18.4%; Score 63; DB 21; Length 225;
 Best Local Similarity 38.9%; Pred. No. 1.7;
 Matches 14; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 7 SRILERSLSRSRVIANVALVPFHPDPDGPVTLTAD 42
 :||| | | : ||. |::|| | | :
 Db 47 trylehqlrslagtylnylgppfnepdnprrgae 82

Search completed: March 7, 2002, 22:05:21
 Job time: 4639 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	175	51.2		5175	2	T20992	hypothetical prote
2	175	51.2		5198	2	T43290	hemiscientific precurs
3	71.5	20.9		1292	2	T48007	P-glycoprotein hom
4	70.5	20.6		1286	2	T02187	probable ABC trans
5	65	19.0		269	2	S53116	hypothetical prote
6	63.5	18.6		237	1	R5D02	ribosomal protein f
7	62.5	18.3		843	2	A27131	epidermal growth f
8	62.5	18.3		1330	1	G0FFE	epidermal growth f
9	62	18.1		1229	2	D85023	P-glycoprotein-lik
10	62	18.1		1229	2	T52319	P-glycoprotein-lik
11	61	17.8		1230	2	E85023	probable P-glycopr
12	60.5	17.7		67	2	D70929	probable ferredoxi
13	60	17.5		211	2	T07594	probable inorganic
14	60	17.5		216	2	T01945	inorganic pyrophos
15	60	17.5		230	2	C86141	protein t25K16.5 [
16	59.5	17.4		178	2	F69716	transcription regu
17	59	17.3		214	2	T01370	inorganic pyrophos
18	59	17.3		552	2	E69438	probable fatty-aci
19	58.5	17.1		275	2	S51658	light harvesting c
20	58.5	17.1		323	2	B48067	ethanolamine-phosp
21	58.5	17.1		415	2	B85436	hypothetical prote
22	58.5	17.1		512	2	T43405	probable dihydroli
23	58	17.0		337	2	S75270	hypothetical prote
24	58	17.0		342	2	D49348	hypothetical prote
25	58	17.0		380	2	C83719	succinoglycan bios
26	57	16.7		114	2	T50971	hypothetical prote
27	57	16.7		283	2	T47174	hypothetical prote
28	57	16.7		359	2	T30382	hypothetical prote
29	57	16.7		399	2	B70349	hypothetical prote

A:Accession: S59316
A:Molecule type: DNA
A:Residues: 1-269
A:Cross-references: EMBL:X91258; NID:g995686; PIDN:CAA62639.1; PID:g995690
A:Experimental source: strain S288C
R:Verhasselt, P.; Voet, M.; Volckaert, G.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64943
A:Accession: S64965
A:Molecule type: DNA
A:Residues: 1-269 <VER>
A:Cross-references: EMBL:Z73300; NID:gl360536; PIDN:CAA97697.1; PID:e245567; PID:gl360536
A:Experimental source: strain S288C
R:Delius, H.; Hebling, U.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64967
A:Accession: S64970
A:Molecule type: DNA
A:Residues: 1-269 <DEW>
A:Cross-references: EMBL:Z73300; NID:gl360536; PIDN:CAA97697.1; PID:e245567; PID:gl360536
A:Experimental source: strain S288C
R:Verhasselt, P.; Volckaert, G.
submitted to the EMBL Data Library, September 1995
A:Reference number: S69393
A:Accession: S69415
A:Molecule type: DNA
A:Residues: 1-269 <VEW>
A:Cross-references: EMBL:X89514; NID:gl297019; PIDN:CAA61706.1; PID:e198756; PID:gl297040
C:Genetics:
A:Map position: 12R
A:Introns: 1/3

```

Query Match      19.0%; Score 65; DB 2; Length 369;
Best Local Similarity 32.7%; Pred. No. 1.6;
Matches 16; Conservative 10; Mismatches 19; Indels 4; Caps 1;

Qy 10 LERSLSRSRVIANVALVPFHDPDG---PVTLTADPVVFQRELQLY 54
      | | | | | | | | | | | | | | | | | | | | |
Db 31 VSRKYLGRNHNWNIYNALNDYDKIGTFTDEVSVTAHPVVPKELTOVF 79

```

RESULT 6
RSD02
ribosomal protein L8.e - slime mold (Dictyostellium discoideum)
N:Alternate names: ribosomal protein DL2; ribosomal protein V1
C:Species: Dictyostellium discoideum
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 22-Jun-1999
C:Accession: S06087
R:Singletton, C.K.
Nucleic Acids Res. 17, 7989, 1989
A:Title: Nucleotide sequence of V1, a ribosomal protein gene from Dictyostellium discoideum
A:Reference number: S06087; MUID:90016885
A:Accession: S06087
A:Molecule type: mRNA
A:Residues: 1-237 <SIN>
A:Cross-references: EMBL:X15710; NID:q7354; PIDN:CAA33741.1; PID:q7355
A:Note: the authors translated the codon CAC for residues 22 and 23 as Gly and TTC for
R:Singletton, C.K.; Delude, R.L.; McPherson, C.E.
Dev. Biol. 119, 433-441, 1987
A:Title: Characterization of genes which are deactivated upon the onset of development
A:Reference number: S09166; MUID:87106347
A:Contents: annotation; differential expression
C:Genetics:
A:Introns: 6/2
C:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: protein biosynthesis; ribosome

Query Match	18.6%	Score 63.5;	DB 1;	Length 237;
Best Local Similarity	36.2%;	Pred. NO. 2.1;		
Matches 21; Conservative	6;	Mismatches 22;	Indels 9;	Gaps 3;

```

Qy      18  SRVTANYALVPFHDPDTPGVTVTLADPVPVFQRELURQL-----YVGGG--DCPEMSVG 67
       :|   ||| :||| ||| :|   |   |   ||| :|   |
Db     128  ARCSGNVATVSHNPDEG-VTLQIITIRPKERSSLRAMIGIVAGGGRIDKPMLKAG 184

RESULT    7
A27131 .
C:epidermal growth factor receptor - fruit fly (Drosophila melanogaster) (
C:Species: Drosophila melanogaster
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 23-May-19
C:Accession: A27131
R:Schejter, E.D.; Segal, D.; Glazer, L.; Shilo, B.Z.
Cell 46, 1091-1101, 1986
A>Title: Alternative 5' exons and tissue-specific expression of the Droso
A:Reference number: A27131; MUID:87002474
A:Accession: A27131
A:Molecule type: mRNA
A:Residues: 1-843 <SCH>
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBN0003731
C:Superfamily: epidermal growth factor receptor; protein kinase homolog
C:Keywords: Atpp; growth factor receptor

```

```
Query Match      18.3%; Score 62.5; DB 2; Length 843;
Best Local Similarity 34.0%; Pred.No.13;
Matches 18; Conservative 8; Mismatches 14; Indels 13; Gaps 2;
```

RESULT 8

Q0FFE
epidermal growth factor receptor - fruit fly (*Drosophila melanogaster*)
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erDB
C:Species: *Drosophila melanogaster*
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C:Accession: A00640; A38021
R:Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B. Z.
Cell 40, 599-607, 1985
A:Title: The *Drosophila* EGF receptor gene homolog: conservation of both hormone
A:Reference number: A00640; MUID:85124611
A:Accession: A00640
A:Molecule type: DNA
A:Residues: 1-1330 <LIV>
A:Cross-references: EMBL:K03054
R:Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A:Title: A *Drosophila* genomic sequence with homology to human epidermal growth
A:Reference number: A38021; MUID:85137938

A: Molecule type: DNA
A: Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK' <WAD>
A: Cross-references: EMBL:X02293; NID:97922; PIDN:CAA26157.1; PID:g929565
C:Comment: This sequence is tentative because the introns have not been identified.
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBgn0003731
A:Map position: 2 57F
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: Arp; autophosphorylation; duplication; glycoprotein; phosphoprotein
F:1-732/Domain: extracellular #status predicted <EXT>
F:733-764/Domain: transmembrane #status predicted <TM>
F:765-1350/Domain: intracellular #status predicted <INT>
F:1808-1072/Domain: protein kinase homology <KIN>
F:1816-824/Region: protein kinase Arp-binding motif
F:122,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #
F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status pre
F:843/Active site: Lys #status predicted
F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status

QY 49 ELRQ - - LYVQGGDCPEMSV 66

RESOLUTION
T07594

probable inorganic pyrophosphatase (EC 3.6.1.1), soluble PPA - potato
C:Species: Solanum tuberosum (potato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 11-May-2000
C:Accession: T07594

R:du Jardin, P.; Rojas-Beltran, J.; Gebhardt, C.; Brasseur, R.
Plant Physiol. 109, 853-860, 1995
A:Title: Molecular cloning and characterization of a soluble inorganic pyrophosphatase
A:Reference number: 216037; MUID:96163189

A:Accession: T07594
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-211 <DUJ>

A:Cross-references: EMBL:Z36894; NID:g534915; PIDN:CAA85362.1; PID:g534916

C:Genetics:

A:Gene: ppa

C:Function:

A:Description: catalyzes the hydrolysis of pyrophosphate

C:Superfamily: inorganic pyrophosphatase

C:Keywords: hydrolase

Query Match 17.5%; Score 60; DB 2; Length 211;
Best Local Similarity 46.3%; Pred. No. 5.1;
Matches 19; Conservative 4; Mismatches 12; Indels 6; Gaps 2;

QY 6 ASRLERSLSRSRVIANVALVPFHDPIGPVTLTADPVWF 46

DB 14 APRLNERILSSLSR--RSVAHPWHDLEIGPEA----PSVF 48

RESULT 14

T01946

inorganic pyrophosphatase homolog Fl104.12 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 22-Jun-1999

C:Accession: T01946

R:Abu-Threideh, J.; Stoneking, T.; Langston, Y.; Trevasakis, E.

submitted to the EMBL Data Library, October 1998

A:Description: The sequence of A. thaliana Fl104.

A:Reference number: Z14466

A:Accession: T01946

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-216 <ABU>

A:Cross-references: EMBL:AF096370; NID:g3695372; PIDN:AAC62786.1; PID:g3695383

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 46/1; 55/3; 77/3; 113/3; 135/3; 176/1; 203/2

A:Note: Fl104.12

C:Superfamily: inorganic pyrophosphatase

Query Match 17.5%; Score 60; DB 2; Length 216;
Best Local Similarity 43.6%; Pred. No. 5.2;
Matches 17; Conservative 6; Mismatches 10; Indels 6; Gaps 2;

QY 8 RILERSLSRSRVIANVALVPFHDPIGPVTLTADPVWF 46

DB 21 RLNERILSSLSK--RSVAHPWHDLEIGP----GAPVIF 53

RESULT 15

C86141

protein T25K16.5 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C86141

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huijzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Ma
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; T
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C86141

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-230 <STO>

A:Cross-references: GB:AE005172; NID:g6715648; PIDN:AAF26475.1; GSPDB:GN00141

C:Genetics:

A:Gene: T25K16.5

A:Map position: 1

Query Match 17.5%; Score 60; DB 2; Length 230;
Best Local Similarity 51.6%; Pred. No. 5.6;
Matches 16; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

QY 6 ASRLERSLSRSRVIANVALVPFHDPIGP 36

DB 15 APRLNERILSSLSR--RSVAHPWHDLEIGP 43

Search completed: March 7, 2002, 22:06:22

Job time: 3535 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 22:13:04 ; Search time 26.04 Seconds
(without alignments)
94.337 Million cell updates/sec

Title: US-09-665-728-1_COPY_56_122

perfect score: 342

Sequence: 1 QVIDGASRILERSLSRSRV.....RELRLYVQGGDCPEMSVG 67

Scoring table: BLOSUM62

decreasing cubic. DEOS0002
 Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

File	Size	Number of hits	Satisfying chosen parameters	100059
...

Minimum DB seq length: 0

Maximum DB seq length: 200000000
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

rose processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt 39:★

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	63.5	18.6	237	1	RL2_DIGDI	P13023 dictyosteli
2	62.5	18.3	1426	1	EGFR_DROME	P04412 drosophila
3	61.5	18.0	287	1	HUPH_BRAJA	P48340 bradyrhizob
4	60.5	17.7	257	1	EC8H_MYCLE	Q07137 mycobacteri
5	60	17.5	211	1	IPVR_SOLTU	Q43187 solanum tub
6	59.5	17.4	178	1	IPST_BACSU	P37554 bacillus su
7	59	17.3	214	1	IPVR_MAIZE	O48556 zea mays (m
8	58.5	17.1	323	1	ECT1_YEAST	P33412 saccharomyc
9	58	17.0	342	1	EXOU_RHIME	P33700 rhizobium m
10	58	17.0	380	1	GRK_BACHD	Q929p2 bacillus ha
11	57.5	16.8	361	1	FDH_PICAN	P33677 pichia angu
12	57	16.7	399	1	Y545_AQUAE	O66822 aquifex aeo
13	57	16.7	484	1	CO1B_MOUSE	Q9wum3 mus musculu
14	57	16.7	484	1	CO1B_RAT	O89046 rattus norv
15	56.5	16.5	1094	1	EX5B_MYCTU	P96920 mycobacteri
16	56	16.4	285	1	YK03_MYCTU	Q10853 mycobacteri
17	55.5	16.2	348	1	ARSL_HUMAN	Q43681 homo sapien
18	55.5	16.2	348	1	ARSL_MOUSE	O54984 mus musculu
19	55.5	16.2	360	1	MK14_CANFA	O02812 canis fami
20	55.5	16.2	360	1	MK14_HUMAN	Q16539 h mitogen-a
21	55.5	16.2	360	1	MK14_MOUSE	P47811 mus musculu
22	55.5	16.2	456	1	GSHR_HAEIN	P43783 haemophilus
23	55.5	16.2	1434	1	VG65_HSVII	Q00106 ictalurid h
24	55	16.1	288	1	RS3A_ANOGA	P52813 anopheles g
25	55	16.1	405	1	CPXK_SACER	P33271 saccharopol
26	55	16.1	469	1	VL2_BPVI	P03109 bovine bapi
27	55	16.1	933	1	RCSC_ECOLI	P14376 escherichia
28	55	16.1	1005	1	DPOL_VARY	P33793 variola vir
29	55	16.1	1006	1	DPOL_VACCC	P20509 vaccinia vi
30	55	16.1	1006	1	DPOL_VACCV	P08856 vaccinia vi
31	54.5	15.9	312	1	M1AA_STRCO	O69967 streptomyce
32	54	15.8	362	1	CKRA_MOUSE	Q9j121 mus musculu
33	54	15.8	424	1	THIK_HUMAN	P09110 homo sapien

34	54	15.8	443	1	W70T_RAT	O35828	rattus norv
35	54	15.8	925	1	W70T_HUMAN	P57737	homo sapien
36	54	15.8	1312	1	P1P1_DROME	P25455	drosophila
37	53.5	15.6	192	1	H541_SOYBN	P30236	glycine max
38	53.5	15.6	552	1	VNS1_BT2VA	P35931	bluetongue
39	53.5	15.6	1093	1	AF17_HUMAN	P55198	homo sapien
40	53	15.5	214	1	IPYR_ORYSA	O22537	oryza sativ
41	53	15.5	215	1	IPYR_HORVD	O23979	hordeum vul
42	53	15.5	298	1	SPEE_SCHPO	Q09741	schizosacch
43	53	15.5	302	1	SPEE_MOUSE	P19623	homo sapien
44	53	15.5	302	1	SPEE_MOUSE	Q64674	mus musculu
45	53	15.5	376	1	WECB_ECOLI	P27828	escherichia

ALIGNMENTS

```

RESULT 1
RL2_DICDI STANDARD; PRT; 237 AA.
ID RL2_DICDI
AC P13023;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S RIBOSOMAL PROTEIN L2.
DN RL2 OR RPGA OR V1.
GN Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=90016885; Pubmed=2678007;
RA Singleton C.K.;
RT "Nucleotide sequence of V1, a ribosomal protein gene from
  Dictyostelium discoideum."
RL Nucleic Acids Res. 17:7989-7989(1989).
CC -1- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; X15710; CAA33741.1; -.
CC PIR; S06087; R5DO2.
CC Dictydb; DD01019; rpgA.
CC InterPro; IPR002171; Ribosomal_L2.
CC Pfam; PF00181; Ribosomal_L2; 1.
CC PROSITE; PS00467; RIBOSOMAL_L2; 1.
CC Ribosomal protein.
KW Ribosomal protein.
SQ SEQUENCE 237 AA; 26298 MW; 77B736C767E35C71 CRC64;

Query Match 18.6%; Score 63.5; DB 1; Length 237;
Best Local Similarity 36.2%; Pred. No. 1;
Matches 21; Conservative 6; Mismatches 22; Indels 9; Gaps 3;

QY 18 SRVTANVALVFFHPDPTGPTVTLTADPVVQFRELRLQL-----YVQGGG--DCPEMSVG 67
: | | | | : | | | | | : | : | : | | | | | : |
Db 128 ARCSGNVATIVSHNPDEG-VFLYQITTRKEERSLARMTGIVAGGGRIDPKMLKAG 184

RESULT 2
EGFR_DROME
ID EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P01868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

```

DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112) (EGFR)
 DE (GURKEN RECEPTOR) (TORPEDO PROTEIN) (DROSOPHILA RELATIVE OF ERBB).
 GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RN SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
 RX MEDLINE=94350209; PubMed=8070664;
 RA Clifford R., Schubach T.;
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
 RT that several genetically defined classes of alleles cluster in
 RT subdomains of the receptor protein.";
 RN Genetics 137:531-550(1994).
 RN [2]
 RN REVISIONS.
 RA Clifford R., Schubach T.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=85124611; PubMed=2982499;
 RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
 RT "The Drosophila EGF receptor gene homolog: Conservation of both
 RT hormone binding and kinase domains.";
 RL Cell 40:599-607(1985).
 RN [4]
 RN SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
 RC STRAIN=OREGON-R; TISSUE=Embryo;
 RX MEDLINE=87002474; PubMed=3093080;
 RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
 RT "Alternative 5' exons and tissue-specific expression of the
 RT Drosophila EGF receptor homolog transcripts.";
 RL Cell 46:1091-1101(1986).
 RN [5]
 RN SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II), AND CHARACTERIZATION.
 RX MEDLINE=99102120; PubMed=9882502;
 RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
 RT "Several levels of EGF receptor signaling during photoreceptor
 RT specification in wild-type, Ellipse, and null mutant Drosophila.";
 RL Dev. Biol. 205:129-144(1999).
 RN [6]
 RN SEQUENCE FROM N.A. (ISOFORM TYPE I).
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Hoyt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle G., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu J., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush L., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [7]
 RN SEQUENCE OF 959-1078 FROM N.A.
 RC STRAIN=DAEKWANYEONG;
 RX MEDLINE=85137938; PubMed=2983232;
 RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
 RT "A Drosophila genomic sequence with homology to human epidermal
 RT growth factor receptor.";
 RL Nature 314:178-180(1985).
 RN [8]
 RN PARTIAL SEQUENCE FROM N.A., AND MUTATION ANALYSIS.
 RX MEDLINE=92038942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.-Z.;
 RT "Interallelic complementation among DER/flb alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";
 RN Genetics 129:191-201(1991).
 RN [9]
 RN REVIEW.
 RX MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.;
 RT "There must be 50 ways to rule the signal: the case of the Drosophila
 RL Cell 89:13-16(1997).
 CC -|- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS.
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOTEROSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -|- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -|- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -|- ALTERNATIVE PRODUCTS: 3 ISOFORMS: TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -|- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW. LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GLANDS.
 CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF052754; AAC08536.1; -
 CC EMBL; AF052753; AAC08536.1; JOINED.
 CC EMBL; AF052754; AAC08535.1; -
 CC EMBL; AF052752; AAC08535.1; JOINED.
 CC EMBL; K03054; AAA51462.1; -
 CC EMBL; K03417; AAA51460.1; -.

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Y13803; CAA74130.1; -;
DR EMBL; AL583925; CAC31918.1; -;
DR HSP; P14604; 2DUB.
DR Leproma; ML2402; -;
DR InterPro; IPR001753; Enoyl-CoA_hydratase.
DR Pfam; PF00378; ECH; 1.
DR PROSITE; PS00166; Enoyl-CoA_HYDRATASE; 1.
KW Fatty acid metabolism; Lyase; Complete proteome.
SQ SEQUENCE 257 AA; 27516 MW; 25B801EB69BCD00 CRC64;

Query Match 17.7%; Score 60.5; DB 1; Length 257;
Best Local Similarity 27.0%; Pred. No. 2.7;
Matches 20; Conservative 7; Mismatches 28; Indels 19; Gaps 2;

QY 2 VIDGASRI-----LERSLSRSRVIANVALVPFHPDIPGVTLTADPVVQREL 51
DB 7 LVGGQVRGVIITLNRQALNALNSQMMNEITNAKELDIDPDVGAILITGSPKV- 61
QY 52 QLYVQGGDCPEMS 65
DB 62 ----AAGADIKEMA 71

RESULT 5
ID IPYR_SOLTU STANDARD; PRT; 211 AA.
AC Q43187;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SOLUBLE INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE PHOSPHO-
DE HYDROLASE) (PPASE).
GN PPA.

OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=96163189; PubMed=8552717;
RA du Jardin P., Rojas-Beltran J., Gebhardt C., Brasseur R.;
RT "Molecular cloning and characterization of a soluble inorganic
pyrophosphatase in potato."
RL Plant Physiol. 109:853-860(1995).
CC -1- CATALYTIC ACTIVITY: PYROPHOSPHATE + H(2)O = 2 ORTHOPHOSPHATE.
CC -1- COFACTOR: MAGNESIUM DEPENDENT.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE PPASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; Z36894; CAA85362.1; -;
DR HSP; P38576; 2PRD.
DR InterPro; IPR001596; Pyrophosphatase.
DR Pfam; PF00719; Pyrophosphatase; 1.
DR ProDom; PD002014; Pyrophosphatase; 1.
DR PROSITE; PS00387; PPASE; 1.
KW Hydrolase; Magnesium.
FT ACT_SITE 61
SQ SEQUENCE 211 AA; 24261 MW; D06115FC6F2AC22A CRC64;

BY SIMILARITY.

Query Match 17.5%; Score 60; DB 1; Length 211;
Best Local Similarity 46.3%; Pred. No. 2.5;
Matches 19; Conservative 4; Mismatches 12; Indels 6; Gaps 2;

QY 6 ASRILERSLSRSRVIANVALVPFHPDIPGVTLTADPVV 46
DB 14 APRLNERILSSISR--RSVAAPHWDLEIGPEA----PSVF 48

RESULT 6
ID SPST_BACSU STANDARD; PRT; 178 AA.
AC P37554;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE STAGE V SPOULATION PROTEIN T.
GN SPOVT.

OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin."
RL DNA Res. 1:1-14(1994).
RN [2]

RP FUNCTION.
RX MEDLINE=96345614; PubMed=8755877;
RA Bayan I., Hobot J., Cutting S.M.;
RT "A compartmentalized regulator of developmental gene expression in
RT Bacillus subtilis."
RL J. Bacteriol. 178:4500-4507(1996).

CC -1- FUNCTION: POSITIVE AND NEGATIVE TRANSCRIPTIONAL REGULATOR OF SIGMA
CC G-DEPENDENT GENES. MAY PROVIDE A MECHANISM OF FEEDBACK CONTROL
CC THAT IS IMPORTANT FOR FORESPORE DEVELOPMENT.

CC -1- SIMILARITY: STRONG, TO B.SUBTILIS ABRB AND ABH. SOME, TO
CC M.TUBERCULOSIS RV2595.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; D26185; BAA05291.1; -;
CC EMBL; Z99104; CAB11832.1; -;
DR Subtilist; BG10119; spoVT.
KW Transcription regulation; DNA-binding; Repressor; Activator;
KW Sporulation; Complete proteome.
SQ SEQUENCE 178 AA; 19742 MW; 19418022DD371180 CRC64;

Query Match 17.48%; Score 59.5; DB 1; Length 178;
Best Local Similarity 32.1%; Pred. No. 2.4;
Matches 18; Conservative 10; Mismatches 15; Indels 13; Gaps 3;

QY 7 SRILERSLSRSRVIANVA---LVPFHPD-----IGPVTLTADP---VVFORE 49
DB 99 SEMLEMTDQRSSVLESDAKSVQLVNGIDEDMNSYTVGPVANGDPGAVVIFSKD 154

RESULT 7
ID IPYR_MAIZE STANDARD; PRT; 214 AA.
AC O48556;


```

DR PROSITE: PS00065; D_2_HYDROXYACID_DH.1; 1.
DR PROSITE: PS00670; D_2_HYDROXYACID_DH.2; 1.
DR PROSITE: PS00671; D_2_HYDROXYACID_DH.3; 1.
KW Oxidoreductase; NAD.
FT INIT_MET 0 0 PROBABLE.
FT NP_BIND 164 199 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 257 257 SUBSTRATE-BINDING (BY SIMILARITY).
SQ SEQUENCE 361 AA; 39779 MW; 0FC001366F9E479B CRC64;

Query Match
Best Local Similarity 26.8%; Score 57.5; DB 1; Length 361;
Matches 15; Conservative 14; Mismatches 18; Indels 9; Gaps 2;

QY 4 DGASRIERSLSRSRVIANVALVPFHPDIPGVTITADVPVVFQRELRLQLYVGGG 59
Db 47 EGQSVLEKNISDADVIIS-----TPFH-----PAYITKERIDRAKKLLVWAGV 93

RESULT 12
Y545_AQUAE
ID Y545_AQUAE STANDARD; PRT; 399 AA.
AC 066822;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AQ_545.
GN AQ_545.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
ON NCBI_TaxId=63363;
RX STRAIN=VF5;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358(1998).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000695; AAC06792.1;
CC KW Hypothetical protein; Complete proteome.
CC SEQUENCE 399 AA; 46953 MW; AA57D8680674318B CRC64;

Query Match
Best Local Similarity 30.2%; Score 57; DB 1; Length 399;
Matches 13; Conservative 11; Mismatches 9; Indels 10; Gaps 2;

QY 4 DGASRIERSLSRSRVIANVALVPFHPDIPGVTITADVPVVF 46
Db 150 DG-EKVMKALVDVRKVLADYE-----GPVVITEDKVIY 182

RESULT 13
COIL_MOUSE
ID COIL_MOUSE STANDARD; PRT; 484 AA.
AC Q9NUM3; Q9CVA2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CORONIN 1B (CORONIN 2).
GN CORO1B.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449467; PubMed=9778037;
RA Okumura M., Kung C., Wong S., Rodgers M., Thomas M.L.;
RT "Definition of family of coronin-related proteins conserved between
RT humans and mice: close genetic linkage between coronin-2 and CD45-
RT associated protein.";
RL DNA Cell Biol. 17:779-787(1998).
RN [2]
RP SEQUENCE OF 163-484 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Stomach;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: MAY BE INVOLVED IN CYTOKINESIS, MOTILITY, AND SIGNAL
CC TRANSDUCTION (BY SIMILARITY).
CC -!- SUBUNIT: BINDS ACTIN (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF143956; AAD32704.1;
CC KW MGD; AK008947; BAB25985.1;
CC MGD; MGI:1345963; Coro1b.
CC InterPro: IPR001680; WD40.
CC Pfam: PF00400; WD40; 3.
CC SMART: SM00320; WD40; 3.
CC PROSITE: PS00678; WD_REPEATS_1; 1.
CC PROSITE: PS50082; WD_REPEATS_2; 2.
CC PROSITE: PS50294; WD_REPEATS_REGION; 1.
CC KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 80 120 WD 1.
FT REPEAT 130 170 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 217 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 444 482 COILED COIL (POTENTIAL).
FT CONFLICT 393 393 R -> G (IN REF. 2).
SQ SEQUENCE 484 AA; 53912 MW; 9631CC02E7EAC72F CRC64;

Query Match
Best Local Similarity 16.7%; Score 57; DB 1; Length 484;
Best Local Similarity 31.4%; Pred. No. 15;

```

Matches 16; Conservative 6; Mismatches 13; Indels 16; Gaps 1;

QY 10 LERSLSRSRVIANVLPFHPDIDGIVTLTADPVVVFQRELRLYVQGGD 60
|| : : : ||:||||| | | | | |
Db 255 LEEPMAQLQELSSNGALLPFYDPTSVV-----YVCGKGD 289

RESULT 14

COLB_RAT STANDARD; PRT; 484 AA.
AC 089046.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CORONIN 1B (CORONIN 2).
GN CORO1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Male D.K., Politopoulou G.;
RT "A cell surface form of rat coronin";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RC -1- FUNCTION: MAY BE INVOLVED IN CYTOKINESIS, MOTILITY, AND SIGNAL
TRANSDUCTION (BY SIMILARITY).
CC -1- SUBUNIT: BINDS ACTIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ006064; CAA06836.1; -
DR InterPro: IPR001680; WD40.
DR Pfam: PF004400; WD40; 3.
DR SMART: SM003320; WD40; 3.
DR PROSITE: PS00678; WD_REPEATS.1; 1.
DR PROSITE: PS00682; WD_REPEATS.2; 2.
DR PROSITE: PS0294; WD_REPEATS_REGION.1.
KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 80 120 WD 1.
FT REPEAT 130 170 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 217 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 447 481 COILED COIL (POTENTIAL).
SQ SEQUENCE 484 AA; 53845 MW; 746EDF119680EBBF CRC64;

Query Match 16.7%; Score 57; DB 1; Length 484;
Best Local Similarity 29.4%; Pred. No. 15;
Matches 20; Conservative 6; Mismatches 14; Indels 28; Gaps 2;

QY 5 GASRLERSL-----SSRSRVIANVLPFHPDIDGIVTLTADPVVVFQRELRLQ 52
| ||: || | : : : ||:||||| | | | | |
Db 238 GFSRMSERQLAWDPENFEEPMAQLQELSSNGALLPFYDPTSVV----- 282
QY 53 LYVQGGD 60
|| | | | |
Db 283 -YVCGKGD 289

RESULT 15
EX5B_MYCTU STANDARD; PRT; 1094 AA.
ID EX5B_MYCTU
AC P96920;

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXODEXVRIBONUCLEASE V BETA CHAIN (EC 3.1.11.5).
GN RECB OR RV0630C OR MT0658 OR MTCY20H10.11C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RC -1- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP
(BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: EXONUCLEOLYTIC CLEAVAGE (IN THE PRESENCE OF
ATP) IN EITHER 5'-TO 3'-OR 3'-TO 5'-DIRECTION TO YIELD 5'-
PHOSPHOOLIGONUCLEOTIDES.
CC -1- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
(BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UVRO SUBFAMILY OF HELICASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z92772; CAB07119.1; -
DR EMBL: AE006361; AAK44882.1; -
DR HSSP: P56255; 2PJR.
DR TIGR: MT0658; -
DR Tuberculist: RV0630c; -
DR InterPro: IPR000212; UvrD-helicase.
DR Pfam: PF00580; UvrD-helicase; 1.
KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
KW DNA repair; Complete proteome.
FT NP_BIND 21 28 ATP (POTENTIAL).
SQ SEQUENCE 1094 AA; 118721 MW; 31262D376875C201 CRC64;

Query Match 16.5%; Score 56.5; DB 1; Length 1094;
Best Local Similarity 33.3%; Pred. No. 45;
Matches 17; Conservative 9; Mismatches 16; Indels 9; Gaps 3;
QY 25 ALVPFHPDIDGP--VTLTADPVVVFQRELRLYVQ---GGGD----CPMSV 66

Db 882 ALLPMHDTPLGPAAAAALTRQIGVDRDLRELDFFEMPLAGGDLRGSRPDSL 932

Search completed: March 7, 2002, 22:13:06
Job time: 459 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 22:07:47 ; Search time 65.89 Seconds
(without alignments)
148.736 Million cell updates/sec

Title: US-09-665-728-1_COPY_56_122

Perfect score: 342

Sequence: 1 QVIDGASRILERSLSRSRV.....RELRLYVGGGDCPMSVG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL17.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mmc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	175	51.2	5198	5 076518	076518 caenorhabdi
2	131.5	38.5	873	11 0921Q8	Q921Q8 mus musc
3	131.5	38.3	891	11 09JHAB	Q9JHAB mus musc
4	131	38.3	536	4 09UMHP	Q9UMHP homo sapien
5	131	38.3	852	4 09Y334	Q9Y334 homo sapien
6	71.5	20.9	1292	10 09M1Q9	Q9M1Q9 arabidopsis
7	70.5	20.6	1286	10 080725	080725 arabidopsis
8	65.5	19.2	584	2 09XCV7	09XCV7 streptomyce
9	65	19.0	269	3 012395	012395 saccharomyc
10	63	18.4	225	4 09UBD9	09UBD9 homo sapien
11	63	18.4	225	11 09QZM3	Q9QZM3 mus musc
12	63	18.4	744	2 006693	Q06693 vibrio mimi
13	62.5	18.3	687	2 09ABL3	Q9ABL3 caulobacter
14	62.5	18.3	3487	2 09FBU0	09FBU0 microcystis
15	62	18.1	1229	10 049749	Q49749 arabidopsis
16	62	18.1	1229	10 09SYI2	Q9SYI2 arabidopsis
17	61.5	18.0	1392	12 09QTC5	Q9QTC5 gallid herp
18	61	17.8	802	3 09HE60	09HE60 neurospora
19	61	17.8	1230	10 09SYI3	Q9SYI3 arabidopsis

20	60.5	17.7	67	2 053937	053937 mycobacteri
21	60.5	17.7	285	4 099773	099773 homo sapien
22	60.5	17.7	487	12 09QQM9	09QQM9 pseudorabdi
23	60.5	17.7	669	5 09VR40	09VR40 drosophila
24	60	17.5	216	10 082597	082597 arabidopsis
25	60	17.5	230	10 09MAM9	09MAM9 arabidopsis
26	60	17.5	290	11 09DL18	09DL18 mus musc
27	59.5	17.4	380	3 093982	093982 zygocacchar
28	59.5	17.4	385	5 09NA54	09NA54 caenorhabdi
29	59.5	17.4	443	4 09HAL4	09HAL4 homo sapien
30	59.5	17.4	702	4 09HOK9	09HOK9 homo sapien
31	59	17.3	130	5 09G252	09G252 tainarys so
32	59	17.3	215	10 09SW10	09SW10 populus tre
33	59	17.3	365	2 09KH85	09KH85 streptomyce
34	59	17.3	552	1 028762	028762 archaeoglob
35	58.5	17.1	275	10 040771	040771 picea abies
36	58.5	17.1	354	3 09UQV5	09UQV5 ustilago ma
37	58.5	17.1	373	10 09FUA4	09FUA4 arabidopsis
38	58.5	17.1	415	10 023192	023192 arabidopsis
39	58	17.0	237	2 073158	073158 synechocyst
40	58	17.0	511	4 09H401	09H401 homo sapien
41	57.5	16.8	1820	5 09VVG0	09VVG0 drosophila
42	57	16.7	114	3 09P3Q9	09P3Q9 neurospora
43	57	16.7	192	2 09RMP5	09RMP5 xenorhabdus
44	57	16.7	283	4 09NSK3	09NSK3 homo sapien
45	57	16.7	359	12 09YMT9	09YMT9 lymantria d

ALIGNMENTS

RESULT 1

076518	076518	PRELIMINARY;	PRT; 5198 AA.
ID	076518	Q10036;	
AC	076518	01-NOV-1998 (TREMREL. 08, Created)	
DT	01-NOV-1998	(TREMREL. 08, Last sequence update)	
DT	01-JUN-2001	(TREMREL. 17, Last annotation update)	
DE	HEMICENTIN PRECURSOR.		
GN	HM-4 OR F15G9.4.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
NCBI_TaxID=6239;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-BRISTOL N2;		
RA	Vogel B.E., Hedgecock E.M.;		
RT	"Hemicentin is Required for Hemidesmosome Mediated Cell Adhesion and Germ-Line Chromosome Segregation in C. elegans."		
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-BRISTOL N2;		
RA	Sulston J., Kershaw J.;		
RL	Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.		
CC	-!- ALTERNATIVE PRODUCTS: TWO FORMS: ISOFORM F15G9.4A AND ISOFORM F15G9.4B (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.		
CC	-!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX DOMAIN.		
CC	EMBL; AF074901; AAC26792.1; -		
DR	EMBL; Z47068; CAA87336.1; -		
DR	EMBL; Z47070; CAA87336.1; JOINED.		
DR	EMBL; Z47070; CAA87345.1; -		
DR	EMBL; Z47068; CAA87345.1; JOINED.		
DR	EMBL; Z47068; CAA87335.1; -		
DR	EMBL; Z47070; CAA87335.1; JOINED.		
DR	EMBL; Z47070; CAA87344.1; -		
DR	EMBL; Z47068; CAA87344.1; JOINED.		
DR	HSSP; P35555; LEWN.		
DR	WormPep; F15G9.4A; CE18595.		
DR	WormPep; F15G9.4B; CE18596.		
DR	InterPro; IPR000152; Asx_hydroxyl.		

DR InterPro: IPR000515; BPD_transp.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR003606; Ig_MHC.
 DR Pfam: PF00047; Ig_47.
 DR SMART: SM00179; EGF_Ca; 1.
 DR SMART: SM00001; EGF_Like; 1.
 DR SMART: SM00408; IGC2; 44.
 DR SMART: SM00410; IGC2; 44.
 DR SMART: SM00327; VWA; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 DR PROSITE: PS01187; EGF_Ca; 2.
 KW Alternative splicing; Calcium-binding; EGF-like domain; Glycoprotein;
 FT Repeat; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 5198 HEMICENTIN.
 FT VARSPIC 1462 5198 MISSING (IN ISOFORM F15G9.4A).
 SQ SEQUENCE 5198 AA; 570816 MW; DA8511FFB2B58D37B CRC64;

Query Match 51.2%; Score 175; DB 5; Length 5198;
 Best Local Similarity 45.5%; Pred. No. 6.6e-12;
 Matches 30; Conservative 16; Mismatches 20; Indels 0; Gaps 0;
 QY 1 QVTDGASRLERSLSRSRVIANVLPFHPDIPGVTLTADPVVFORLQLYVGGGD 60
 DB 53 QVREGAAKIFKTVMAOREKLIYINVPFHPDIPGLIINTDSTYFMKLSKVYVHGGD 112
 QY 61 CPMSV 66
 DB 113 CPKTL 118

RESULT 2
 ID Q921Q8 PRELIMINARY; PRT; 873 AA.
 AC Q921Q8;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 94.9 KDA PROTEIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RA Rowen L., Madan A., Qin S., Shaffer T., Ratcliffe A., Abbasi N.,
 RA Dickhoff R., James R., Loretz C., Lasky S., Hood L.;
 RT "Sequence of the mouse major histocompatibility locus class III
 RT region.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF109905; AAC84152.1;
 DR InterPro: IPR002035; VWFA.
 DR SMART: SM00327; VWA; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 873 AA; 94894 MW; 5761E3620F8D831C CRC64;

Query Match 38.5%; Score 131.5; DB 11; Length 873;
 Best Local Similarity 47.5%; Pred. No. 1.7e-07;
 Matches 28; Conservative 9; Mismatches 21; Indels 1; Gaps 1;

QY 6 ASRLERSLSRSRVIANVLPFHPDIPGVTLTADPVVFORLQLYVGGGD 64
 DB 346 ARRIVEQRQGSMPFVF-YILVFPHPDIPGVFTTSDPSFQKLNEIHALGGDEPM 403

RESULT 3
 ID Q9JHA8 PRELIMINARY; PRT; 891 AA.
 AC Q9JHA8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE G7C (G7C PROTEIN).
 GN G7C.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/RLJ, AND BALB/C; TISSUE=BRAIN;
 RX MEDLINE=98211706; PubMed=9551980;
 RA Snoek M., Teuscher C., van Vugt H.;
 RT "Molecular analysis of the major MHC recombinational hot spot located
 RT within the G7c gene of the murine class III region that is involved in
 RT disease susceptibility.";
 RL J. Immunol. 160:266-272(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/RLJ, AND BALB/C; TISSUE=BRAIN;
 RX MEDLINE=20260998; PubMed=10803853;
 RA Snoek M., Albertella M.R., van Kooij M., Wixon J., van Vugt H.,
 RA de Groot K., Campbell R.D.;
 RT "G7c, a novel gene in the mouse and human major histocompatibility
 RT complex class III region, possibly controlling lung tumor
 RT susceptibility.";
 RL Immunogenetics 51:383-386(2000).
 DR EMBL; AF134318; AAF69177.1;
 DR EMBL; AF134319; AAF61401.1;
 DR InterPro: IPR002035; VWFA.
 DR SMART: SM00327; VWA; 1.
 SQ SEQUENCE 891 AA; 95983 MW; 1EA26B5121118720 CRC64;

Query Match 38.5%; Score 131.5; DB 11; Length 891;
 Best Local Similarity 47.5%; Pred. No. 1.7e-07;
 Matches 28; Conservative 9; Mismatches 21; Indels 1; Gaps 1;

QY 6 ASRLERSLSRSRVIANVLPFHPDIPGVTLTADPVVFORLQLYVGGGD 64
 DB 338 ARRIVEQRQGSMPFVF-YILVFPHPDIPGVFTTSDPSFQKLNEIHALGGDEPM 395

RESULT 4
 ID Q9UMP9 PRELIMINARY; PRT; 536 AA.
 AC Q9UMP9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE G7C PROTEIN (FRAGMENT).
 GN G7C.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Aguado B.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Albertella M.R., Campbell R.D.;
 RT "Characterisation of the novel gene G7c located in the class III
 RT region of the human Major Histocompatibility Complex.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]

09 OCT 2008 (ATTN:MMET-LS), BASE SEQUENCE UPDATE)

DR InterPro: IPR001140: ABC_transporter_tmern.
DR InterPro: IPR003439: ABC_transporter.
DR InterPro: IPR001687: ATP_GTP_A.
DR Pfam: IPR001525: C5_DNA_meth.
DR Pfam: PF00664: ABC_membrane; 2.
DR Pfam: PF00005: ABC_tran; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR PROSITE: PS00095; C5_MTASE_2; UNKNOWN_1.
KW ATP-binding; Transport.
SQ SEQUENCE 1286 AA; 139027 MW; 97D2ABAFBA698E6 CRC64;

Query Match 20.6%; Score 70.5; DB 10; Length 1286;
Best Local Similarity 27.8%; Pred. No. 7.9;
Matches 22; Conservative 11; Mismatches 23; Indels 23; Gaps 3;

QY 9 ILERSLSRSRVIANVLPVPHDPDIPVTL-----TADPVVFOR 48
DB 1075 LVGESGSGKSTVIA--LQRFYDPSGETILDGVEIKSLRLKWLKQQTGLVSVQEPILFNE 1132
QY 49 ELRLYVOG-GGDCPEMSV 66
DB 1133 TIRANAYKGGDAESEI 1151

RESULT 8
Q9XCV7 PRELIMINARY; PRT; 584 AA.
AC Q9XCV7;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE JADJ.
OS Streptomyces venezuelae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=54571;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISP5230;
RA Kulowski K., Wendt-Pienkowski E., Han L., Yang K., Vining L.C.,
RA Hutchinson C.R.;
RT "functional characterization of the jadI gene as a cyclase that forms
RT anugcyclicinones.";
RL J. Am. Chem. Soc. 0:0-0(1999).
CC -1- COFACTOR: BIOTIN (BY SIMILARITY).
DR EMBL: AF126429; RAD37851.1; -;
DR HSSP: P24182; IBNC.
DR InterPro: IPR001882; Biotin.
DR InterPro: IPR000089; Biotin_lipoyl.
DR InterPro: IPR000901; CPase.
DR Pfam: PF00364; biotin_lipoyl; 1.
DR Pfam: PF00289; CPase_Lchain; 1.
DR PROSITE: PS00188; BIOTIN; 1.
DR PROSITE: PS00867; CPASE_2; UNKNOWN_1.
KW Biotin.
SQ SEQUENCE 584 AA; 61926 MW; 4384224B32ADC64B CRC64;

Query Match 19.2%; Score 65.5; DB 2; Length 584;
Best Local Similarity 27.8%; Pred. No. 13;
Matches 22; Conservative 11; Mismatches 19; Indels 27; Gaps 5;

QY 2 VIDGASRLERSLSRSRVIANVLPVPHDPDIPVTLADP----- 43
DB 387 IVTGASR--EQALQRAARVAERVEGMATAIPFHQAVVTDPD-----FTADPFRRVHTRW 439
QY 44 --VVFORELRQLYVGGGD 60
DB 440 IETEFVNEIKPFAPAGEAD 458

RESULT 9
Q12395 PRELIMINARY; PRT; 269 AA.
AC Q12395;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE ORF YLR128W.
GN YLR128W OR L3111 OR L9233.2.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FY23/RD005;
RA Verhasselt P., Volckaert G.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Delius H.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Delius H., Hebling U.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Pauley A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favellio A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA Taich A., Travaaskis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: X89514; CAA61706.1; -;
DR EMBL: X91258; CAA62639.1; -;
DR EMBL: Z73308; CAA97697.1; -;
DR EMBL: U53877; AAB82374.1; -;
DR SGD: S0004118; YLR128W.
SQ SEQUENCE 269 AA; 32203 MW; C8D829C941466180 CRC64;

Query Match 19.0%; Score 65; DB 3; Length 269;
Best Local Similarity 32.7%; Pred. No. 5.8;
Matches 16; Conservative 10; Mismatches 19; Indels 4; Gaps 1;

QY 10 LERSLSRSRVIANVLPVPHDPDIPG----PVTLPADPVVFORELRQLY 54
DB 31 VSKRYLQRNHNINVALNDYDKETGTFDDEVSTVAHPVPKELTQVF 79
RESULT 10
Q9UBD9 PRELIMINARY; PRT; 225 AA.
ID Q9UBD9

AC Q9UBD9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE NEUROTROPHIN-1/B-CELL STIMULATING FACTOR-3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432254; PubMed=10500198;
RA Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lille J., Scully S.,
RA Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,
RA Simonet W.S., Boone T., Chang M.-S.;
RT "Novel neurotrophin-1/B cell-stimulating factor-3: A cytokine of the
RT IL-6 family";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99382254; PubMed=10448081;
RA Shi Y., Wang W., Yourey P.A., Gohari S., Zukauskas D., Zhang J.,
RA Ruben S., Alderson R.F.;
RT "Computational EST database analysis identifies a novel member of the
RT neurotrophic cytokine family";
RL Biochem. Biophys. Res. Commun. 262:132-138(1999).
DR EMBL; AF176912; AAD00992.1; -;
DR EMBL; AF172854; AAD54284.1; -;
DR EMBL; AF176911; AAF00981.1; -;
SQ SEQUENCE 225 AA; 25176 MW; E2DD4B6280833B55 CRC64;

Query Match 18.4%; Score 63; DB 4; Length 225;
Best Local Similarity 38.9%; Pred. No. 8.2;
Matches 14; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 7 SRILERSLSRSRVIANVALVPHDPDIPGVTLTAD 42
Db 47 TRYLEHQLRSLAGTYLNLVGLPFPNPDNPPRLGAE 82

RESULT 11
Q9QZM3
ID Q9QZM3 PRELIMINARY; PRT; 225 AA.
AC Q9QZM3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NEUROTROPHIN-1/B-CELL STIMULATING FACTOR-3.
GN BSF3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432254; PubMed=10500198;
RA Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lille J., Scully S.,
RA Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,
RA Simonet W.S., Boone T., Chang M.-S.;
RT "Novel neurotrophin-1/B cell-stimulating factor-3: A cytokine of the
RT IL-6 family";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).
DR EMBL; AF176913; AAF00993.1; -;
DR MGD; MGI:1930088; Bsf3.
SQ SEQUENCE 225 AA; 25261 MW; 68B1FEAAB7F1A950 CRC64;

Query Match 18.4%; Score 63; DB 11; Length 225;
Best Local Similarity 38.9%; Pred. No. 8.2;
Matches 14; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 7 SRILERSLSRSRVIANVALVPHDPDIPGVTLTAD 42

Db 47 TRYLEHQLRSLAGTYLNLVGLPFPNPDNPPRLGAE 82

RESULT 12
Q06693
ID Q06693 PRELIMINARY; PRT; 744 AA.
AC Q06693;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HEAT-LABILE HEMOLYSIN.
GN VMHA.
OS Vibrio mimicus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=674;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33653;
RX MEDLINE=97274002; PubMed=9128176;
RA Kim G.T., Lee J.Y., Huh S.H., Yu J.H., Kong I.S.;
RT "Nucleotide sequence of the vmhA gene encoding hemolysin from Vibrio
RT mimicus";
RL Biochim. Biophys. Acta 1360:102-104(1997).
DR EMBL; U68271; AAB58399.1; -;
DR InterPro; IPR000772; Ricin_B_lectin.
DR InterPro; IPR001340; Hemolysn_pore.
DR Pfam; PF00652; Ricin_B_lectin; 1.
DR SMART; SMO0458; RICIN; 1.
SQ SEQUENCE 744 AA; 83087 MW; E7E71F4FC8B83F89 CRC64;

Query Match 18.4%; Score 63; DB 2; Length 744;
Best Local Similarity 39.7%; Pred. No. 34;
Matches 23; Conservative 6; Mismatches 19; Indels 10; Gaps 4;

QY 19 RVIANVALVPHDPDIPGVTLTADPVVFOR--ELRQLYVQGG-----GDCPMSVG 67
Db 600 RTITSYTNV-FHEQESSVGLGLTQGVNQVRKDNQLYVRAGNAIDALGTAPELLVG 656

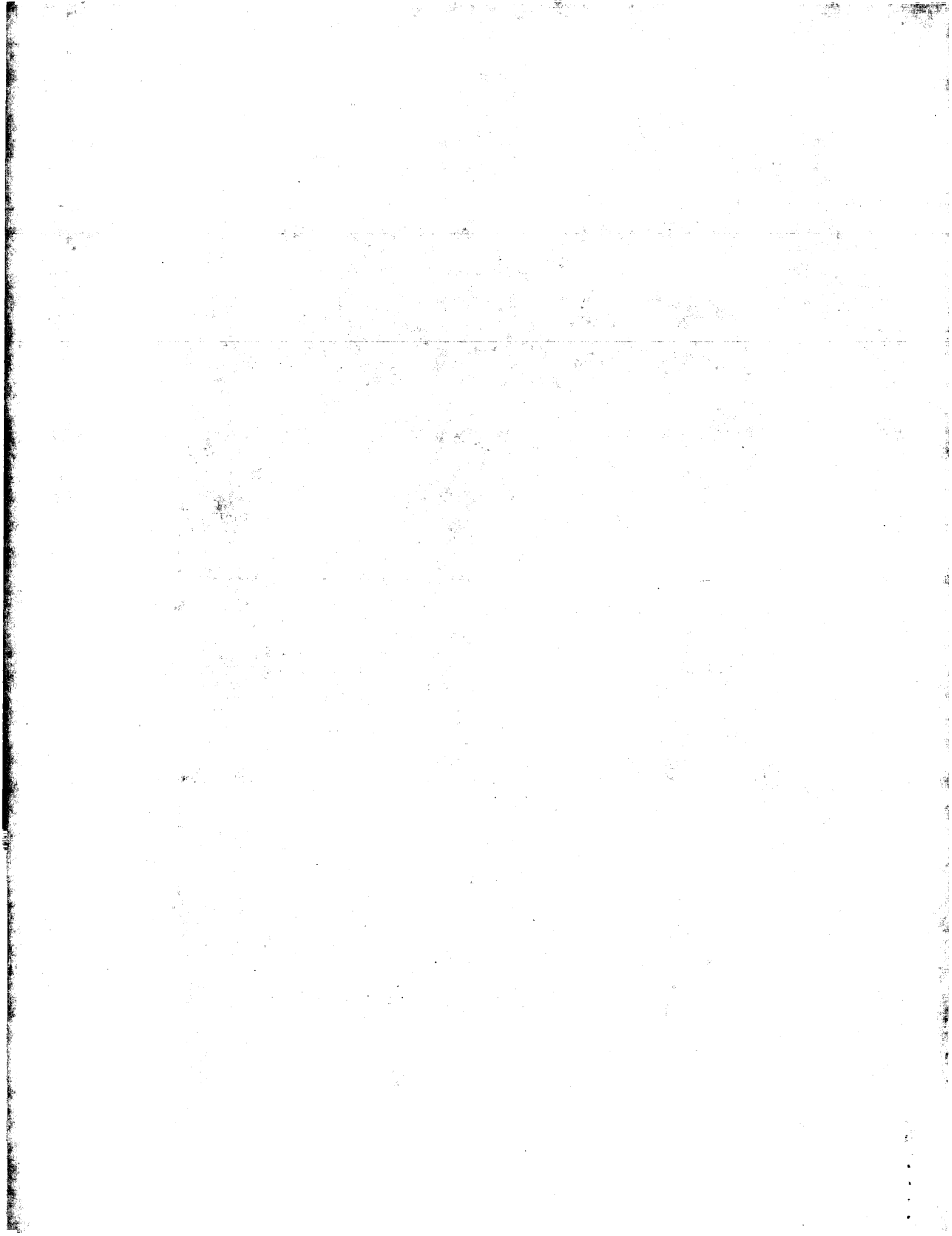
RESULT 13
Q9ABL3
ID Q9ABL3 PRELIMINARY; PRT; 687 AA.
AC Q9ABL3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE TONB-DEPENDENT RECEPTOR.
GN CC0214.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=11259647; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ullnerback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005695; AAK22201.1; -;
DR TIGR; CC0214; -;
KW Receptor; Complete proteome.
SQ SEQUENCE 687 AA; 72786 MW; 29415A726A3BF753 CRC64;

Query Match 18.3%; Score 62.5; DB 2; Length 687;

049749 ID 049749 PRELIMINARY: PRT: 1229 AA.

Search completed: March 7, 2002, 22:07:49
Job time: 3167 sec

Search completed: March 7, 2002, 22:07:49
Job time: 3167 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 20:01:47 ; Search time 148.15 seconds
(without alignments)
5966.266 Million cell updates/sec

Title: US-09-665-728-2
Perfect score: 1031
Sequence: 1 tctagcgaacccttcgccc.....agcccggtcgcgcgcgc 1031

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1031	100.0	1031	22	AAF82464
2	51.6	5.0	2004	18	AAF82464
3	50.4	4.9	2338	19	AAF23249
4	50.4	4.9	2338	21	AAF38195
5	47.2	4.6	2338	12	AAF41483
6	43.4	4.2	2745	22	AAF09568
7	41.8	4.1	9063	22	AAF23688
8	41.2	4.0	1260	19	AAF41730
9	41	4.0	1747	20	AAF82459
10	40.8	4.0	1160	20	AAF25208
11	40.4	3.9	2271	15	AAF062176

12	40.4	3.9	2271	15	AAF05334
13	40.2	3.9	2214	19	AAF22682
14	40.2	3.9	3331	19	AAF22683
15	40.2	3.9	3331	20	AAF232021
16	40.2	3.9	3331	22	AAF09078
17	39.6	3.8	2502	16	AAF080752
18	39.6	3.8	2502	19	AAF65782
19	39.6	3.8	2502	20	AAF63403
20	39.4	3.8	4748	22	AAF06573
21	38.8	3.8	1929	19	AAF57472
22	38.8	3.8	2668	21	AAF00337
23	38.8	3.8	12152	22	AAF08699
24	38.6	3.7	1598	14	AAF04061
25	38.4	3.7	2461	17	AAF41544
26	38.2	3.7	1995	12	AAF41484
27	38.2	3.7	1995	19	AAF23250
28	38.2	3.7	1995	21	AAF38196
29	38	3.7	1588	22	AAF26304
30	37.6	3.6	567	21	AAF29550
31	37.6	3.6	71989	21	AAF29349
32	37.4	3.6	571	20	AAF89296
33	37.4	3.6	3120	21	AAF12510
34	37.4	3.6	5145	21	AAF98196
35	37.4	3.6	6085	18	AAF70153
36	37.4	3.6	31422	21	AAF92302
37	37.4	3.6	47981	22	AAF30757
38	37.2	3.6	355	21	AAF74463
39	37.2	3.6	993	21	AAF30765
40	37.2	3.6	993	21	AAF30778
41	37.2	3.6	1616	20	AAF33961
42	37.2	3.6	2260	19	AAF41995
43	37.2	3.6	5335	21	AAF66076
44	37.2	3.6	6225	20	AAF55273
45	37.2	3.6	6225	21	AAF20843

ALIGNMENTS

RESULT 1
AAF82464
ID AAF82464 standard; cDNA; 1031 BP.
XX
AC AAF82464;
XX
DT 29-JUN-2001 (first entry)
XX
Rat cDNA clone P00210D09.
XX
Rat; secreted factor; P00210D09; cardiatic; nephrotropic;
KW antiinflammatory; gene therapy; cardiac disease; renal disease;
KW inflammatory disease; ss.
XX
Rattus norvegicus.
OS
XX
FH Key Location/Qualifiers
FT CDS 96..923
FT
FT /*tag= a
FT /product= "Rat secreted factor"
FT sig_peptide 96..158
FT /*tag= b
FT mat_peptide 159..923
FT /*tag= c
FT
FT WO200123419-A2.
XX
PN 05-APR-2001.
XX
PD
XX
PF 27-SEP-2000; 2000WO-US26582.
XX
PR 27-SEP-1999; 99US-0156277.
XX
PA (SCIO-) SCIOS INC.

XX Stanton LW, Kapoun AM;
 XX WPI; 2001-328177/34.
 DR P-PSDB; AAB83147.
 XX Novel secreted factor encoded by clone P00210B09 useful for diagnosing,
 PT treating and/or preventing various cardiac, renal and inflammatory
 PT diseases -
 XX
 XX Claim 1; Fig 2; 69pp; English.
 PS
 XX The present sequence encodes a novel secreted factor. The invention
 CC relates to a polypeptide comprising a sequence of at least 80% identity
 CC to residues 22-122 of a 275 amino acid sequence, or a sequence encoded
 CC by a nucleic acid hybridising under stringent conditions to the
 CC complement of the coding region comprising 1031 nucleotides, and having
 CC at least one biological activity of the polypeptide encoded by clone
 CC P00210B09. The polypeptides and polynucleotides of the invention are
 CC useful for the treatment of cardiac, renal and inflammatory diseases.
 CC The polynucleotides are useful in antisense mediated gene inhibition
 CC and in gene therapy. The polypeptides are useful in assays for
 CC identifying lead compounds that may be used as therapeutic agents in
 CC the treatment of cardiac, kidney or inflammatory diseases.
 XX
 XX Sequence 1031 BP; 215 A; 317 C; 308 G; 191 T; 0 other;
 SQ

Query Match 100.0%; Score 1031; DB 22; Length 1031;
 Best Local Similarity 100.0%; Pred. No. 7.7e-237;
 Matches 1031; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tctagcaaaccttcgcccgttagagcagactgactgacgtccactctatccctcgacct 60
 Db 1 tctagcaaaccttcgcccgttagagcagactgactgacgtccactctatccctcgacct 60

Qy 61 gcgcgtccattagggctgagcctccggtcagctgacgtcagccttagggcgagctctgc 120
 Db 61 gcgcgtccattagggctgagcctccggtcagctgacgtcagccttagggcgagctctgc 120

Qy 121 cgctgctccctggcagctacacagtagtgccggtgagcgtccatctgtagagccacga 180
 Db 121 cgctgctccctggcagctacacagtagtgccggtgagcgtccatctgtagagccacga 180

Qy 181 agacgtgtcccccgcacagagagacccacctggccttcgtcttcgtatgtaccggct 240
 Db 181 agacgtgtcccccgcacagagagacccacctggccttcgtcttcgtatgtaccggct 240

Qy 241 ccattgtgggagctctgtagcagctgacgtgacgtcagcgtcagcgttcttgagcgagtc 300
 Db 241 ccattgtgggagctctgtagcagctgacgtgacgtcagcgtcagcgttcttgagcgagtc 300

Qy 301 tgagcagccgagcgggtcagcgtcagcgtcagcgtcagcgtcagcgtcagcgtcagcgtc 360
 Db 301 tgagcagccgagcgggtcagcgtcagcgtcagcgtcagcgtcagcgtcagcgtcagcgtc 360

Qy 361 ttggccacagtgacccctcagcgtcagcgtcagcgtcagcgtcagcgtcagcgtcagcgtc 420
 Db 361 ttggccacagtgacccctcagcgtcagcgtcagcgtcagcgtcagcgtcagcgtcagcgtc 420

Qy 421 atgttcaggaggtggtgtagctgcccagaaatagtgctgggggcccataaggtgccttgg 480
 Db 421 atgttcaggaggtggtgtagctgcccagaaatagtgctgggggcccataaggtgccttgg 480

Qy 481 aggttgcgaaccccggtctctcatctacgtctctcgtgagtcggtcgcagagactacc 540
 Db 481 aggttgcgaaccccggtctctcatctacgtctctcgtgagtcggtcgcagagactacc 540

Qy 541 acaagaagaatagctcctcagctcctcagctgagcagctgagcagctgagcagctgagcagct 600
 Db 541 acaagaagaatagctcctcagctcctcagctgagcagctgagcagctgagcagctgagcagct 600

Qy 601 tgactggggagctcggtgacccgacccacctggctacgtggtcttttgaggagatcgct 660
 Db 601 tgactggggagctcggtgacccgacccacctggctacgtggtcttttgaggagatcgct 660

Db 601 tgactggggagctcggtgacccgacccacctggctacgtggtcttttgaggagatcgct 660
 Qy 661 ccaccagttctggccaagtgttcagctggaagaagcagcagcagcagcagcagcagcagcagc 720
 Db 661 ccaccagttctggccaagtgttcagctggaagaagcagcagcagcagcagcagcagcagcagc 720
 Qy 721 ggggtggagtcgcgcacccatccagcgtcccaaaagtctcctgctcagcagcagcagcagcagc 780
 Db 721 ggggtggagtcgcgcacccatccagcgtcccaaaagtctcctgctcagcagcagcagcagcagc 780
 Qy 781 agggcgaaacacacatggagaatcccttttgaccccccagcttgaagaagtcaccatctcac 840
 Db 781 agggcgaaacacacatggagaatcccttttgaccccccagcttgaagaagtcaccatctcac 840
 Qy 841 tgagcgggccaagcgcctgagatcgaagtccgggagccacccactgggtatgtcccaagggttcac 900
 Db 841 tgagcgggccaagcgcctgagatcgaagtccgggagccacccactgggtatgtcccaagggttcac 900
 Qy 901 ctctctctctgacgaagactgagctggaagccagcgtgagcgtgagcgtgagcgtgagcgtgagc 960
 Db 901 ctctctctctgacgaagactgagctggaagccagcgtgagcgtgagcgtgagcgtgagcgtgagc 960
 Qy 961 tgaggagatggctcagcgaataaaatgtctgctcctcacacacacacacacacacacacacacac 1020
 Db 961 tgaggagatggctcagcgaataaaatgtctgctcctcacacacacacacacacacacacacacac 1020

Qy 1021 cgagcgccgc 1031
 Db 1021 cgagcgccgc 1031

RESULT 2
 AAT85356
 ID AAT85356 standard; DNA; 2004 BP.
 XX
 XX AAT85356;
 XX AC
 XX 09-DEC-1997 (first entry)
 XX
 XX Nephila clavipes spider silk protein 2 Kb DNA sequence.
 XX
 XX High strength film; fibre; woven article; parachutes; sails;
 KW absorber; body armour; heavy metal; biological weapon; chemical;
 KW flavour; fragrance; Nephila clavipes; ss.
 XX
 XX Nephila clavipes.
 XX
 XX Key Location/Qualifiers
 FH CDS 40..1980
 FT /*tag= a
 FT /product= Silk_protein
 XX
 XX WO9708315-A1.
 XX
 XX 06-MAR-1997.
 XX
 XX 22-AUG-1996; 96WO-US13767.
 XX
 XX 22-AUG-1995; 95US-0517694.
 XX
 XX (BASE/) BASEL R M.
 PA (ELIO/) ELION G R.
 XX
 XX Basel RM, Elion GR;
 XX
 XX WPI; 1997-179272/16.
 DR P-PSDB; AAW27178.
 XX
 XX New opt. multimerised DNA sequences encoding spider silk protein -
 PT contg. both repetitive and non-repetitive sequences, useful for
 PT making high strength films, fibres, woven articles etc.
 XX

PD 23-NOV-1999.
 XX
 PF 04-OCT-1994; 94US-0317844.
 XX
 PR 15-APR-1991; 91US-0684819.
 PR 20-APR-1990; 90US-0511792.
 XX
 PA (UYWY-) UNIV WYOMING.
 XX
 PI Hinman MB, Xu M, Lewis RV;
 XX WPI; 2000-061225/05.
 DR P-PSDB; AAY59070.
 XX
 PT Isolated DNA, vector and transformed cell encoding for and useful in
 PT the production of spider silk protein
 XX
 PS Example 3; Fig 6A-D; 65pp; English.
 XX
 CC The invention provides isolated cDNA molecules coding for spider silk
 CC proteins. The spider silk proteins are characterized by repeating alpha
 CC and beta regions and optional variable regions. The DNA sequences are
 CC useful in the production of spider silk protein by recombinant DNA
 CC techniques. The recombinant spider silk proteins may be used for the
 CC production of fibers. The present sequence represents the cDNA encoding
 CC the spider silk protein 1, derived from the major ampullate gland of
 CC Nephila clavipes.
 XX
 SQ Sequence 2338 BP; 564 A; 432 C; 918 G; 424 T; 0 other;

Query Match 4.9%; Score 50.4; DB 21; Length 2338;
 Best Local Similarity 49.3%; Pred. No. 0.01;
 Matches 132; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
 QY 514 tctcgatgcccgcggtgctctgctgacgaggaatgagctctgagctctctgcagc 573
 DB 459 tcttgaagcagggtgcagagcaggtgattgagtgagcagaggtgcagcagc 518
 QY 574 tgaagcagtcgaggtgctctgctgacgaggaactgcggtgacgcacccaccctg 633
 DB 519 agcagcagccgaggtgctggaacagcgatagcgtgcttctggaacaggtgcgg 578
 QY 634 gctacctggtttgagagagatgcctccacagttctggaagttctccagctggaca 693
 DB 579 acaaggaggtatggagacttggaagccaaggtgctgacgagggaggtattagtgaca 638
 QY 694 agcagcaggtgctgaggtgttaaaagtgggtgagtcctccagcctccaaagttc 753
 DB 639 aggtgcaggtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 698
 QY 754 atctgctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 781
 DB 699 acaaggtgctggaacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 726

RESULT 5
 AAQ14183
 ID AAQ14183 standard; cDNA; 2338 BP.
 XX
 AC AAQ14183;
 XX
 DT 15-JAN-1992 (first entry)
 XX
 DE N.clavipes dragline silk protein coding sequence.
 XX
 KW protein superfiber; major ampullate silk; orb web spider; ss.
 XX
 OS Nephila clavipes.
 XX
 FT Key Location/Qualifiers
 FT CDS 1..2157
 FT /*tag= a

FT
 XX
 PN EP452925-A.
 XX
 PD 23-OCT-1991.
 XX
 PF 18-APR-1991; 91EP-0106217.
 XX
 PR 20-APR-1990; 90US-0511792.
 XX
 PA (UYWY-) UNIV OF WYOMING.
 XX
 PI Lewis RV, Xu M, Hinman M;
 XX WPI; 1991-312199/43.
 DR P-PSDB; AAR14308.
 XX
 PT DNA encoding spider silk protein-1 and 2 and variants - isolated
 PT from Nephila clavipes, for prodn. of spider silk protein and
 PT fibres having desired characteristics
 XX
 PS Claim 4; Page 23; 48pp; English.
 XX
 CC A N.clavipes major ampullate gland cDNA library was screened with
 CC probes based on peptide fragments of the purified spider silk
 CC protein (see AAQ14185). Positive plaques were identified and the
 CC spider silk protein 1 coding sequence was determined. See also
 CC AAQ14184.
 XX
 SQ Sequence 2338 BP; 566 A; 433 C; 916 G; 423 T; 0 other;

Query Match 4.6%; Score 47.2; DB 12; Length 2338;
 Best Local Similarity 48.5%; Pred. No. 0.06;
 Matches 130; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
 QY 514 tctcgatgcccgcggtgctctgctgacgaggaatgagctctgagctctctgcagc 573
 DB 459 tcttgaagcagggtgcagcagcagcagcagcagcagcagcagcagcagcagcagc 518
 QY 574 tgaagcagtcgaggtgctctgctgacgaggaactgcggtgacgcacccaccctg 633
 DB 519 agcagcagccgaggtgctggaacagcgatagcgtgcttctggaacaggtgcgg 578
 QY 634 gctacctggtttgagagagatgcctccacagttctggaagttctccagctggaca 693
 DB 579 acaaggaggtatggagacttggaagccaaggtgctgacgagggaggtattagtgaca 638
 QY 694 agcagcaggtgctgaggtgttaaaagtgggtgagtcctccagcctccaaagttc 753
 DB 639 aggtgcaggtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 698
 QY 754 atctgctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 781
 DB 699 acaaggtgctggaacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 726

RESULT 6
 AAD09568
 ID AAD09568 standard; cDNA; 2745 BP.
 XX
 AC AAD09568;
 XX
 DT 10-SEP-2001 (first entry)
 XX
 DE Human transporter and ion channel-17 (TRICH-17) cDNA.
 XX
 KW Human; transporter and ion channel-17; TRICH-17; cystic fibrosis; mood;
 KW gene therapy; amyotrophic lateral sclerosis; amnesia; muscular dystrophy;
 KW hypertension; angina; neurological disorder; asthma; bipolar disorder;
 KW dementia; depression; Alzheimer's disease; epilepsy; vaccine; arrhythmia;
 KW Pick's disease; ischaemic cerebrovascular disease; AIDS; anxiety; stroke;
 KW Huntington's disease; Parkinson's disease; cerebral neoplasm; allergy;

CC (by viruses, bacteria, fungi, parasites, protozoa or helminths).
 XX Sequence 9063 BP; 2492 A; 2546 C; 2654 G; 1371 T; 0 other;
 SQ

Query Match 4.1%; Score 41.8; DB 22; Length 9063;
 Best Local Similarity 55.0%; Pred. No. 1.7;
 Matches 82; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
 QY 465 atcaaggctgcgtgaggttgccaaacccggctcttcacgtcttcctcgatgcc 524
 DB 7943 acaagtcagtcgagcagcgtttcaacgcccgcaggttcctctcgatgccg 8002
 QY 525 cgtgccaaggactaccacaagaagatgagctcctcgagctcctcgagcagtcg 584
 DB 8003 gatgacaagtagcaccgcatgaagactgctcctcctcagcagcagcagcgcg 8062
 QY 585 caggtggtctgctgactggggactg 613
 DB 8063 gccctgaacgcgtgcagaggatggatg 8091

RESULT 8
 AAV41730
 ID AAV41730 standard; DNA; 1260 BP.
 AC AAV41730;
 XX
 XX 20-NOV-1998 (first entry)
 XX
 DE Codon-optimised Ramy3D signal fused to DNA encoding mature AAT.
 KW Protein expression; monocotyledon plant cell;
 KW glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;
 KW ATIII; human serum albumin; HSA; subtilisin BPN'; treatment; emphysema;
 KW antithrombotic; blood replacement; ss.
 XX Synthetic.
 OS Homo sapiens.

Key Location/Qualifiers
 misc_feature 1..75
 /tag= a
 /note= "codon-optimised Ramy3D signal sequence"
 misc_feature 76..1260
 /tag= b
 /note= "encodes mature AAT"

W09836085-A1.
 XX
 XX 20-AUG-1998.
 XX
 XX 13-FEB-1998; 98WO-US03068.
 XX
 XX 13-FEB-1997; 97US-0038170.
 XX 13-FEB-1997; 97US-0037991.
 XX 13-FEB-1997; 97US-0038168.
 XX 13-FEB-1997; 97US-0038169.
 XX
 XX (PHYT-) APPLIED PHYTOLOGICS INC.
 XX Rodriguez RL, Sutliff TD;
 XX WPI: 1998-467179/40.
 XX
 XX Expressing mature, glycosylated proteins in monocotyledonous plant
 PT cells - from chimeric gene including signal peptide sequence,
 PT specifically therapeutic agents and industrial enzymes
 XX
 XX Disclosure: Pages 34 iii-iv; 53pp; English.
 XX
 XX The present sequence encodes a fusion protein of codon-optimised Ramy3D
 CC signal sequence/mature alpha1-antitrypsin (AAT). The protein is used

CC to exemplify the invention. The specification describes a method for
 CC producing mature heterologous protein in monocotyledonous plant cells.
 CC The method comprises transforming the cells with a chimeric gene
 CC comprising a monocotyledon transcription regulator, inducible either
 CC during seed maturation or by adding/removing a small molecule, DNA
 CC encoding the heterologous protein, and DNA encoding a signal peptide,
 CC with the signal peptide causing secretion of the protein from the cell.
 CC Proteins expressed in this manner include mature glycosylated alpha
 CC 1-antitrypsin (AAT) with a glycosylation pattern that significantly
 CC increases its serum half-life, mature glycosylated antithrombin III
 CC (ATIII), mature human serum albumin (HSA) having the native folding
 CC pattern as shown by bilirubin-binding characteristics, or mature active
 CC subtilisin BPN'. These proteins are useful therapeutically (e.g. AAT for
 CC treating emphysema, ATIII as antithrombotic and HSA as blood replacement)
 CC or as industrial enzymes (BPN' is used in detergents).
 XX
 SQ Sequence 1260 BP; 287 A; 428 C; 350 G; 195 T; 0 other;

Query Match 4.0%; Score 41.2; DB 19; Length 1260;
 Best Local Similarity 47.6%; Pred. No. 1.4;
 Matches 121; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
 QY 464 catcaaggctgcgtgaggttgccaaacccggctcttcacgtcttcctcgatgcc 523
 DB 759 catccagcactgcagaagctctccagctgggtgctcctcatgaagcactggggaacgc 818
 QY 524 ccgtgccaaaggactaccacaagaagatgagctcctcgagctcctcgagcagtcg 583
 DB 819 caccgcatctcttcctgcgcgagggcagcagctccagcagctggagaacgagctgac 878
 QY 584 gcaggtggtctctgctgactggggactgcggtgagcagcaccacccctggctaccctggc 643
 DB 879 gcagcagcatcatcagaagttcctctggagaacgagcagggcgctccgctagctccaccc 938
 QY 644 ttctgaggagatcgccctccaccagcttctgcccagtgcttcagctgagcagcagcaggt 703
 DB 939 ccggaagctgagcatcaccgagcagctgacgagcagctgagcagcagctgctggcagctg 998
 QY 704 gtcggaggtgttaa 717
 DB 999 caccgaagctctca 1012

RESULT 9
 AAV82459/c
 ID AAV82459 standard; DNA; 1747 BP.
 XX
 XX AAV82459;
 XX
 XX 16-MAR-1999 (first entry)
 XX
 XX Triticum sp. cysteine proteinase #4 encoding DNA.
 DE Triticum; wheat seed; cysteine proteinase; gluten; baking; ds.
 KW Triticum sp.
 OS Triticum sp.
 XX
 XX Key Location/Qualifiers
 FT CDS 79..1497
 FT /*tag= a
 XX
 XX JP10327886-A.
 XX
 XX 15-DEC-1998.
 XX
 XX 27-MAR-1998; 98JP-0098140.
 XX
 XX 31-MAR-1997; 97JP-0114946.
 XX
 XX (SHOS) SHOWA SANGYO CO.
 XX WPI: 1999-109255/10.
 DR

Search completed: March 7, 2002, 21:17:38
Job time: 4551 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 20:44:19 ; Search time 1988.65 seconds
(without alignments)
6843.926 Million cell updates/sec

Title: US-09-665-728-2_COPY_96_920
Perfect score: 825
Sequence: 1 atgacgcctaggcgacgct.....ctcctcttctgatgcaagac 825

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0
Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl : *
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.om.*
20: em.or.*
21: em.ov.*
22: em.pat.*
23: em.ph.*
24: em.pl.*
25: em.ro.*
26: em.sts.*
27: em.sy.*
28: em.un.*
29: em.vi.*
30: em.htgo_hum.*
31: em.htgo_inv.*
32: em.htgo_rod.*
33: em.htg_hum.*
34: em.htg_inv.*
35: em.htg_rod.*
36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

Result No.	Score	Match	Length	DB	ID	Description
1	825	100.0	1031	6	AX108634	AX108634 Sequence
2	336.8	40.8	18207	9	AF156100	AF156100 Homo sapi
3	236.8	28.7	191750	2	AC024113	AC024113 Mus muscu
4	185.2	22.4	15597	3	AF074901	AF074901 Caenorhab
5	177	21.5	182211	9	AL360004	AL360004 Human DNA
6	171.2	20.8	191750	2	AC024113	AC024113 Mus muscu
7	171.2	20.8	230755	2	AC024618	AC024618 Mus muscu
8	108.4	13.1	41345	3	CEFI5G9	Z47088 Caenorhabdi
9	93.6	11.3	1610	9	HS245418	AJ245418 Homo sapi
10	91.4	11.1	4312	10	AF134319	AF134319 Mus muscu
11	89.4	10.8	157029	9	AL135796	AL135796 Human DNA
12	69.8	8.5	87461	9	HS118G19	AL121996 Human DNA
13	69.8	8.5	180248	9	AC023275	AC023275 Homo sapi
14	65.4	7.9	717	11	HS348TG1	AL159308 STS from
15	61.6	7.5	356	11	HS348TG1	Z51502 H.sapiens (
16	51.6	6.3	1873	3	NCU37520	U37520 Nephila cla
17	50.4	6.1	2336	3	NEUPSF	M37137 N.clavipes
18	50.4	6.1	2338	6	AR088543	AR088543 Sequence
19	50.4	6.1	2338	6	I92789	I92789 Sequence 1
20	48.6	5.9	87810	2	AC022987	D85391 Mus musculu
21	47	5.7	4359	10	D85391	D85391 Mus musculu
22	47	5.7	41173	1	SCI41	AY033407 Streptomy
23	46.2	5.6	3867	1	AY033407	AY033407 Myxococcu
24	46	5.6	1026	3	AF350272	AF350272 Streptomy
25	44.8	5.4	43349	1	SCF41	AL117387 Streptomy
26	44.2	5.4	44109	1	SC5F1	AL450165 Streptomy
27	44	5.3	12606	1	AE005995	AE005995 Caulobact
28	43.8	5.3	87810	2	AC022987	AC022987 Homo sapi
29	43.4	5.3	2745	6	AX179756	AX179756 Sequence
30	43	5.2	2047	10	BC003909	BC003909 Mus muscu
31	43	5.2	5991	10	AF199422	AF199422 Mus muscu
32	43	5.2	6423	10	AF199421	AF199421 Mus muscu
33	43	5.2	198677	1	AE001863	AE001863 Deinococc
34	42.6	5.2	132470	9	HS203P18	Z97180 Human DNA s
35	42.2	5.1	100000	3	AP000503	AP000503 Homo sapi
36	42.2	5.1	125350	2	AC020768	AC020768 Homo sapi
37	42.2	5.1	163682	9	HSDJ71117	AL132713 Human DNA
38	42.2	5.1	180283	9	AF134726	AF134726 Homo sapi
39	41.8	5.1	1682	1	STMGLNR	L03213 Streptomyce
40	41.8	5.1	11717	1	AE004509	AE004509 Pseudomon
41	41.8	5.1	32354	1	SCD84	AL353816 Streptomy
42	41.6	5.0	3203	10	AF008561	AF008561 Mus muscu
43	41.6	5.0	12356	10	AF134318	AF134318 Mus muscu
44	41.6	5.0	135545	10	MMHC213L3	AF109905 Mus muscu
45	41.6	5.0	224020	2	AC087117	AC087117 Mus muscu

ALIGNMENTS

RESULT	1
AX108634	
LOCUS	AX108634 1031 bp DNA
DEFINITION	Sequence 2 from Patent WO0123419.
ACCESSION	AX108634
VERSION	AX108634.1 GI:13923866
KEYWORDS	Norway rat.
SOURCE	Rattus norvegicus
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 1031)
AUTHORS	Stanton,L.W. and Kapoun,A.M.
TITLE	Differentially expressed genes
JOURNAL	Patent: WO 0123419-A 2 05-APR-2001;
SCIOS INC. (US)	
FEATURES	Location/Qualifiers
source	1. .1031
/organism="Rattus norvegicus"	

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 191750)
Worley,K.C.
Direct Submission
Submitted (24-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 4, 2001 this sequence version replaced gi:11096415.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MABZ
Center clone name: RP23-277F6
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 74% of reads
Chemistry: Dye-terminator Big Dye; 26% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 159536 bases at least Q40
Consensus quality: 173428 bases at least Q30
Consensus quality: 179682 bases at least Q20
Estimated insert size: 180449; sum-of-contrigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.2x in Q20 bases; sum-of-contrigs estimation

COMMENT

```

* 165460 165559: gap of unknown length
* 165560 168279: contig of 2720 bp in length
* 168280 168379: gap of unknown length
* 168380 170847: contig of 2468 bp in length
* 170848 170947: gap of unknown length
* 170948 171115: contig of 4168 bp in length
* 171116 175215: gap of unknown length
* 175216 177870: contig of 2655 bp in length
* 177871 177970: gap of unknown length
* 177971 181412: contig of 3442 bp in length
* 181413 181512: gap of unknown length
* 181513 183055: contig of 1343 bp in length
* 183056 183155: gap of unknown length
* 183156 184937: contig of 1782 bp in length
* 184938 185037: gap of unknown length
* 185038 187608: contig of 2571 bp in length
* 187609 187708: gap of unknown length
* 187709 189220: contig of 1512 bp in length
* 189221 189320: gap of unknown length
* 189321 190396: contig of 1076 bp in length
* 190397 190496: gap of unknown length
* 190497 191750: contig of 1254 bp in length.

FEATURES             Location/Qualifiers
     source            1..191750
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /chromosome="11"
                        /clone="RP23-277F6"
BASE COUNT   47260 a 47982 c 47655 g 46018 t 2835 others
ORIGIN

Query Match      28.7%; Score 236.8; DB 2; Length 191750;
Best Local Similarity 91.9%; Pred. No. 1.6e-37;
Matches 250; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Oy 1 atgacgcctagggcgagctcctgcgcgtctctctggcgacctacacagtagtgaggcgcg 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191185 ATGAGCGCCTGGCGGCGAGCTCTCGGATGCTCGTGGCGATCTCTACAGCAGTGGCGCG 191244

Oy 61 gcgcacatctgatgagccacgagacgctgtcccgccagagagagagagagagagagag 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191245 GTGGTCAGCTGTGACGGCGCCGCCAGGACGCTGTCCCGCCGACAGGAGAGCCACCCTA 191304

Oy 121 gccttcgcttcgatgtcacccggtccatgtgggacgatctgtgacgtgatgcagtgacgcgc 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191305 GCCTTCGCTTCGATGTACCGGCTCCATGTGGATGACCTGTGAGTGTATCGAGTATCGAGCGC 191364

Oy 181 gcctcagcattctggagcgagctctgagcagcgccgagcgccggtccatcgccaactatgcgc 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191365 GCCTCGCGCATCTCGAGCGCGAGCGCTGAGCAGCGCGGTCATCGCGGTCATCGCCCACTATGCG 191424

Oy 241 ctggtgctttccacgacccagacattggccc 272
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191425 CTGTGCTTTCACAGCCCGAGGTAGAGCCCC 191456

RESULT      4
AF074901
LOCUS
DEFINITION
Caenorhabditis elegans hemicentin precursor (him-4) mRNA, complete
cds.
ACCESSION
AF074901
VERSION
AF074901.1
GI:3328185
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
1 (bases 1 to 15597)
AUTHORS
Vogel,B.E. and Hedgecock,E.M.
TITLE
Hemicentin, a conserved extracellular member of the immunoglobulin
superfamily, organizes epithelial and other cell attachments into

```

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 19880: contig of 19880 bp in length
1 19881 19980: gap of unknown length
1 19981 38620: contig of 18640 bp in length
1 38621 38720: gap of unknown length
1 38721 51091: contig of 12371 bp in length
1 51092 51191: gap of unknown length
1 51192 63586: contig of 12395 bp in length
1 63587 63687: gap of unknown length
1 63688 76765: contig of 13079 bp in length
1 76766 76865: gap of unknown length
1 76866 92384: contig of 15519 bp in length
1 92385 92484: gap of unknown length
1 92485 105755: contig of 13270 bp in length
1 105756 105854: gap of unknown length
1 105855 114346: contig of 8492 bp in length
1 114347 114446: gap of unknown length
1 114447 122760: contig of 8314 bp in length
1 122761 122861: gap of unknown length
1 122862 130074: contig of 7213 bp in length
1 130075 130174: gap of unknown length
1 130175 135344: contig of 5171 bp in length
1 135345 135444: gap of unknown length
1 135445 140753: contig of 5309 bp in length
1 140754 140853: gap of unknown length
1 140854 146305: contig of 5452 bp in length
1 146306 146405: gap of unknown length
1 146406 150873: contig of 4468 bp in length
1 150874 150973: gap of unknown length
1 150974 155410: contig of 4437 bp in length
1 155411 155511: gap of unknown length
1 155512 159685: contig of 4174 bp in length
1 159686 159785: gap of unknown length
1 159786 163180: contig of 3396 bp in length
1 163181 163280: gap of unknown length
1 163281 165459: contig of 2179 bp in length

```


QY 159 tctgatgcaggtgatgcagcggccctcagcattcttgaggcgcagctctgagagccgcgag 218
 Db 150 TTGTGTCAGGTACGGGAAGGTGCTGCTAAAAATCTCAAAATCTGATGGCAACGTGA 209
 QY 219 ccggggtcagcgcacacatgcgctgttccacacccagacattggccagtgac 278
 Db 210 AAGCTAATCTACACTACATCATGCTTCTTTTTCACGATCCATCTTAGGAGAGATTAT 269
 QY 279 cctcagcggcgaccagtggttttcagagagctgagacacactctatgttcaggagg 338
 Db 270 CAATACCACCGACTCAACTACTTTCATGAGACAACTCAGTAAGTCTAGCTGCACGGAGG 329
 QY 339 tggtagctccacagaaatgagtgctgggggccaatcgaagctgcgctggaggttgcacaacc 398
 Db 330 AGGTGATTTCAGAGAACGCTTACCGGTATTCTCAAGGCACCTTCAAAATTTCTTACC 389
 QY 399 cggctcctcatctcagctcttcgagtcgctgacccgctcccaagagactaccacaagaatga 458
 Db 390 GTCCCTCATTTATGTGTTTCTCAGTCATGCCCGATCAAGAGATTACCACTTAGAAGATGA 449
 QY 459 gctcctcagagctcctgcagctgaagcagtcgcaggtggtcttcgtgactggggactg 518
 Db 450 AGTTCTTAATAACAATTCAAGAGAAACAAAGCTCAGTTGTTTTCGTGATGACTGGTGACTG 509
 QY 519 cggtagcgcaccccccctgctactgctgttcttgagagatcgctccacagttctgg 578
 Db 510 TGGTAATGAACACATCCAGGATTCGGTACATATGAAAAAATCGCGCTGCATCTTTTGG 569
 QY 579 ccaagtgtccagctggacacagcagtgctcgaggtgttaagtggttaagtggtgagtcgcg 638
 Db 570 ACAAGTTTCCATCTTGAAACAGTCAGTCAGTACCGTATTAGATATGTCAGACATGC 629
 QY 639 catcaggcctccaaagtctatctgtctcagcagcacccagggaggagggcgcaacacac 698
 Db 630 TGCTAAACAGAGAAGATTCATTGTATGACGAGCAGCTGAACGGGGAGGAACAGTGTGTC 689
 QY 699 atggagaatcccttttgaccacagcttgaaggaagtaccatctcactgagcgg 752
 Db 690 TAGAAACATCCAGTAGACAAAGCATCTCTCAGAGCTCACCATCTCATTGTTCAGG 743

RESULT 5

AL360004 AL360004 182211 bp DNA PRI 16-AUG-2001
 LOCUS Human DNA sequence from clone RP11-88G17 on chromosome 9, complete
 DEFINITION sequence.
 ACCESSION AL360004
 VERSION AL360004.22 GI:15212049
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 AUTHORS Kimberley A.
 TITLE Direct Submission
 JOURNAL Submitted (16-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Aug 18, 2001 this sequence version replaced gi:15020463.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em.; EMBL; Sw.;
 SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr9
 RP11-88G17 is from the library RPCI-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-88G17 The true
 right end of clone RP11-202H3 is at 75398 in this sequence.

FEATURES

source 1..182211
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-88G17"
 /clone_lib="RPCI-11.1"
 /clone="44209"
 /note="Sequence from AC006241 sequenced by WIBR".
 misc_feature 50279..50351
 /note="Sequence from AC006241 sequenced by WIBR"
 misc_feature 93314
 /note="Tandem repeat. Forced join. Gap sized to be
 approximately 150bp from restriction digest data and
 spanning pUC clone."
 BASE COUNT 40368 a 47059 c 48127 g 46657 t
 ORIGIN

Query Match 21.5%; Score 177; DB 9; Length 182211;
 Best Local Similarity 78.2%; Pred. No. 9.3e-26;
 Matches 226; Conservative 0; Mismatches 60; Indels 3; Gaps 1;
 QY 1 atgacgcctagggcgcagctcctgcgctgctctctggcgacctacacagtagtgcgggcg 60
 Db 137309 ATGATGCGGGGGCGCGCTCCTCGCGGTGCTGACCGGGTCTCTGCGGACGTGCGAGTG 137368
 QY 61 gcgggtcacatctgatgacccacgaagacgcgtgtctcccgccacagggagagacccactg 120
 Db 137369 GCAGTGGCC--GGGGCGCGCGGACGGTAATGCCCCCACCACCGGGGACGCCACCTG 137425
 QY 121 gccttcgtcttcgatgcacgcgctccatgtgggacgatctgatgcaggtgatgcagcgc 180
 Db 137426 GCCTTCGTCTTCGACGTCAACGGCTCCATGTGGGACGAACGTGTCAGGTGATCGATGGC 137485
 QY 181 gctcacgcattcttgagcgcagctctgagcgcgcagccgaggttcacgcacacatgcg 240
 Db 137486 GCCTCGCATTTCTGGAACGCGAGTCTGAGCGCGCCGACGAGGCCATCGCCAACTACGCG 137545
 QY 241 ctggtgcctttccacagaccagacattggccccagtgaccctcagcggcg 289
 Db 137546 CTGTTGCCCTTCCACGACACCGAGGTAGCGCCCCCGCCGACCCCGCGCGCG 137594

RESULT 6

AC024113/c AC024113 191750 bp DNA HTG 07-JAN-2001
 LOCUS Mus musculus chromosome 11 clone RP23-277F6, WORKING DRAFT
 DEFINITION SEQUENCE, 29 unordered pieces.
 ACCESSION AC024113
 VERSION AC024113.12 GI:12025591
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 191750)
 AUTHORS Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,

AC024618
AC024618.2 GI:10280861
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
1 (bases 1 to 230755)
Mus musculus, clone RP23-22C1
Unpublished
2 (bases 1 to 230755)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgaiter, B., Brown, A., Burkett, G., Campolano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vasiliiev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission

TITLE
JOURNAL
COMMENT

Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 23, 2000 this sequence version replaced gi:7139711.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5311
Center clone name: 22_C_1

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 206138 bases at least Q40
Consensus quality: 218628 bases at least Q30
Consensus quality: 223560 bases at least Q20
Insert size: 226355; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 45 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2864: contig of 2864 bp in length
* 2865 2964: gap of 100 bp
* 2965 4029: contig of 1065 bp in length
* 4030 4129: gap of 100 bp
* 4130 5611: contig of 1482 bp in length
* 5612 5711: gap of 100 bp
* 5712 6711: contig of 1000 bp in length

6712 6811: gap of 100 bp
* 6812 8073: contig of 1262 bp in length
* 8074 8173: gap of 100 bp
* 8174 9200: contig of 1027 bp in length
* 9201 9300: gap of 100 bp
* 9301 10629: contig of 1329 bp in length
* 10630 10729: gap of 100 bp
* 10730 12013: contig of 1284 bp in length
* 12014 12113: gap of 100 bp
* 12114 14005: contig of 1892 bp in length
* 14006 14105: gap of 100 bp
* 14106 15760: contig of 1655 bp in length
* 15761 15860: gap of 100 bp
* 15861 18390: contig of 2530 bp in length
* 18391 18490: gap of 100 bp
* 18491 20892: contig of 2402 bp in length
* 20893 20992: gap of 100 bp
* 20993 23366: contig of 2374 bp in length
* 23367 23466: gap of 100 bp
* 23467 26126: contig of 2660 bp in length
* 26127 26226: gap of 100 bp
* 26227 29160: contig of 2934 bp in length
* 29161 29260: gap of 100 bp
* 29261 33742: contig of 4482 bp in length
* 33743 33842: gap of 100 bp
* 33843 37207: contig of 3365 bp in length
* 37208 37307: gap of 100 bp
* 37308 41498: contig of 4191 bp in length
* 41499 41598: gap of 100 bp
* 41599 45456: contig of 3858 bp in length
* 45457 45556: gap of 100 bp
* 45557 50797: contig of 5241 bp in length
* 50798 50897: gap of 100 bp
* 50898 54710: contig of 3813 bp in length
* 54711 54810: gap of 100 bp
* 54811 59379: contig of 4569 bp in length
* 59380 59479: gap of 100 bp
* 59480 63217: contig of 3738 bp in length
* 63218 63317: gap of 100 bp
* 63318 66324: contig of 3007 bp in length
* 66325 66424: gap of 100 bp
* 66425 70575: contig of 4151 bp in length
* 70576 70675: gap of 100 bp
* 70676 75597: contig of 4922 bp in length
* 75598 75697: gap of 100 bp
* 75698 80005: contig of 4308 bp in length
* 80006 80105: gap of 100 bp
* 80106 86908: contig of 6803 bp in length
* 86909 87008: gap of 100 bp
* 87009 93267: contig of 6259 bp in length
* 93268 93367: gap of 100 bp
* 93368 97688: contig of 4321 bp in length
* 97689 97788: gap of 100 bp
* 97789 102663: contig of 4875 bp in length
* 102664 102763: gap of 100 bp
* 102764 120018: contig of 17255 bp in length
* 120019 120118: gap of 100 bp
* 120119 125200: contig of 5082 bp in length
* 125201 125300: gap of 100 bp
* 125301 131332: contig of 6032 bp in length
* 131333 131432: gap of 100 bp
* 131433 137931: contig of 6499 bp in length
* 137932 138031: gap of 100 bp
* 138032 146338: contig of 8307 bp in length
* 146339 146438: gap of 100 bp
* 146439 154167: contig of 7729 bp in length
* 154168 154267: gap of 100 bp
* 154268 161318: contig of 7051 bp in length
* 161319 161418: gap of 100 bp
* 161419 169637: contig of 8219 bp in length
* 169638 169737: gap of 100 bp
* 169738 178139: contig of 8402 bp in length
* 178140 178239: gap of 100 bp

* 178240 188969: contig of 10730 bp in length
* 188970 189069: gap of 100 bp
* 189070 198835: contig of 9766 bp in length
* 198836 198935: gap of 100 bp
* 198936 211500: contig of 12565 bp in length
* 211501 211600: gap of 100 bp
* 211601 224426: contig of 12826 bp in length
* 224427 224526: gap of 100 bp
* 224527 230755: contig of 6229 bp in length.

FEATURES
source
1. .230755
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-22C1"
/clone_lib="RPCI-23 Female Mouse BAC"
misc_feature
1. .2864
/note="assembly_fragment"
clone_end:SP6
vector_side:left
2965. .4029
/note="assembly_fragment"
misc_feature
4130. .5611
/note="assembly_fragment"
misc_feature
5712. .6711
/note="assembly_fragment"
misc_feature
6812. .8073
/note="assembly_fragment"
misc_feature
8174. .9200
/note="assembly_fragment"
misc_feature
9301. .10629
/note="assembly_fragment"
misc_feature
10730. .12013
/note="assembly_fragment"
misc_feature
12114. .14005
/note="assembly_fragment"
misc_feature
14106. .15760
/note="assembly_fragment"
misc_feature
15861. .18390
/note="assembly_fragment"
misc_feature
18491. .20892
/note="assembly_fragment"
misc_feature
20993. .22366
/note="assembly_fragment"
misc_feature
23467. .26126
/note="assembly_fragment"
misc_feature
26227. .29160
/note="assembly_fragment"
misc_feature
29261. .33742
/note="assembly_fragment"
misc_feature
33843. .37207
/note="assembly_fragment"

Query Match 20.8%; Score 171.2; DB 2; Length 230755;
Best Local Similarity 90.7%; Pred. No. 1.2e-24;
Matches 194; Conservative 0; Mismatches 18; Indels 2; Gaps 1;
Qy 614 aggtgttaagtgggtgagtcgcgcacatccaggccctccaaagtctcatctgtgtcagcag 673
Db 112066 AGGTGTTAAAGTGGGTGAGTCAGCAATCCAGGCCCTCCAAAGTTTCATCTGCTGTCGAG 112065
Qy 674 accacgaggaggagggcaacacacatcaggaataccctttgaccccgcttgaaggag 733
Db 112066 ACCATGAGGAGGAGGCGGACGACACGCTGGAGATCCCTTTGACCCCGACCTGAGGAGG 112125
Qy 734 tcaccatctcactgagcgggccagggcctgagatcgaagtcgcgggaccactgggtatgt 793
Db 112126 TCACCATTTGCATGACGCGACACGAGGCGCTGAGATTGAAGTCGGGACCCACTGGGTATGCT 112185
Qy 794 --ccagggttcaactctctctctgtatgcaagac 825
Db 112186 GCCCCAGGGCTCACCTCTTCTTTGGATACAAGAC 112219

RESULT 8
CEFI5G9
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .41345
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="F15G9"
join(2873. .3089,3208. .3273,3480. .3823,3869. .3947,
4131. .4264)
/gene="F15G9.la"
join(2873. .3089,3208. .3273,3480. .3823,3869. .3947,
4131. .4264)
/gene="F15G9.la"
/note="CDNA EST yk102el2.3 comes from this gene"
CDNA EST yk102el2.5 comes from this gene
CDNA EST yk168h5.3 comes from this gene
CDNA EST yk168h5.5 comes from this gene
CDNA EST yk450c6.3 comes from this gene
CDNA EST yk450c6.5 comes from this gene
CDNA EST yk268g4.3 comes from this gene
CDNA EST yk268g4.5 comes from this gene
CDNA EST yk252c2.3 comes from this gene
CDNA EST yk252c2.5 comes from this gene
CDNA EST yk527d8.3 comes from this gene
CDNA EST yk558g5.3 comes from this gene

CEFI5G9 41345 bp DNA INV 20-JUN-2001
Caenorhabditis elegans cosmid F15G9, complete sequence.
247068
247068.1 GI:634018
HTG; Fibrinectin type III repeat; Transfer-RNA; tRNA-Ala.
Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
none.
1 (bases 1 to 41345)
Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
The C.elegans Sequencing Consortium.
2 (bases 1 to 41345)
Sulston,J.E.
Direct Submission
Submitted (24-DEC-1994) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone F15G9.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone F15G9 is at 1 in this sequence. The true
right end of clone F15G9 is at 4082 in
sequence 247070.
The true left end of clone T09B9 is at 41245 in this sequence. The
true right end of clone C43C3 is at 8835 in this sequence. The
start of this sequence (1. .101) overlaps with the end of sequence
247067.
The end of this sequence (41245. .41345) overlaps with the start of
sequence 247070.
For a graphical representation of this sequence and its analysis
see:- [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F15G9)
[name=F15G9](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F15G9).

Location/Qualifiers
1. .41345
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="F15G9"
join(2873. .3089,3208. .3273,3480. .3823,3869. .3947,
4131. .4264)
/gene="F15G9.la"
join(2873. .3089,3208. .3273,3480. .3823,3869. .3947,
4131. .4264)
/gene="F15G9.la"
/note="CDNA EST yk102el2.3 comes from this gene"
CDNA EST yk102el2.5 comes from this gene
CDNA EST yk168h5.3 comes from this gene
CDNA EST yk168h5.5 comes from this gene
CDNA EST yk450c6.3 comes from this gene
CDNA EST yk450c6.5 comes from this gene
CDNA EST yk268g4.3 comes from this gene
CDNA EST yk268g4.5 comes from this gene
CDNA EST yk252c2.3 comes from this gene
CDNA EST yk252c2.5 comes from this gene
CDNA EST yk527d8.3 comes from this gene
CDNA EST yk558g5.3 comes from this gene


```
JOURNAL Immunogenetics 51 (4-5), 383-386 (2000)
MEDLINE 20260998
REFERENCE 3 (bases 1 to 4312)
AUTHORS Snoek,M., Van Kooij,A., van Vugt,H. and de Groot,K.E.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1999) Molecular Genetics, the Netherlands Cancer
Institute, Plesmanlaan 121, Amsterdam 1066 CX, The Netherlands
FEATURES
    Location/Qualifiers
        1..4312
            /organism="Mus musculus"
            /strain="BALB/c"
            /db_xref="taxon:10090"
            /chromosome="17"
            /map="19.00 cM; between G7a and G7"
            /tissue_type="brain"
            1..4312
                /gene="G7c"
                /note="D17H6S56E-3; genomic structure of the G7c gene is
                presented in GenBank Accession Number AF134318; similar to
                the human sequence in GenBank Accession Number AJ245418"
                /allele="d"
                205..2880
                    /gene="G7c"
                    /codon_start=1
                    /product="G7c protein"
                    /protein_id="AAF61401.1"
                    /db_xref="GI:7381105"
                    /translation="MLPEVEVPLSHLGLPILLQLLLPPTSAFFPNIWSLLAAGSVT
                    HDITTEALNVTLVLEQPHGRPRLLVEDYRGRTLLADDIFAAYFCPGFSSRRFR
                    AALGSRANAAQDFLPFAKSNPDLPDAERLVGCRTRVLGALRETLVAARALEVTLA
                    RQRLGAALHALQDFVSHSNWVELGEROPHLLMPROELNSLAQVDDPTCSDGSLSC
                    PGNMLDSTLLTSGYFGMHPAPKPKCSHGHHFDQSSQPPRGGINKSDTSPSPHHK
                    LHLQAEVALLASIEAFSLRLGRDKAFSLDLITPASSLSFVLDITSGMGEINAA
                    LQIARRIVEQSGPMPEVFIYLVFHPDGFQPVFTTSDPDSFWQKNEIHALGGDE
                    PEMCLSALEALLPLPSDIFVTDASPKDALLNRYESLTRRRCRVTLVLTEDPS
                    RTGRRRREALSLRFEPEYEAIRASGEVFTTKDOYIQDVAAIVGSMAGLVTLPLD
                    PVFTTGPCEVSDLSLMQVTRMHGDISFWIKSPAGVSGQPEEGIGPLCHTRRF
                    QFNWITDPRTGTWELQVAAAGTPRVRVQAQSLDFLFHFGISVEDGHPGLYPLT
                    QPVAGLQLLVLTGTSRKLKVGQGFQFHVLRVRPEGTQIGRVSLSEVPVPGVR
                    LLAASLPPTLLSVSPFSLVGGGSESLRTPAQPCSVAPVLLLELSGPDFTLPG
                    SKAPLSVIFSFGPDILRTSVNPSFLTSLNSRLRLGINSAMGRWLEVPDSAA
                    PDSVMTVTVAAGGASQVPTTFLRLVLAQSSKQDLDPANSAAPVLPVPSALL
                    PSTLVTCGAGGAGMAGRWCTGVGVFLGLGTSW"
                    1295..1524
                        /misc_feature
                        /gene="G7c"
                    /note="exons around recombinational hot spot; corresponds
                    to sequence presented in GenBank Accession Number
                    AF008561"
                    4254..4259
                        /polyA_signal
                        /gene="G7c"
BASE COUNT 830 a 1310 c 1166 g 1006 t
ORIGIN

Query Match 11.1%; Score 91.4; DB 10; Length 4312;
Best Local Similarity 52.4%; Pred. No. 1.7e-08;
Matches 225; Conservative 0; Mismatches 201; Indels 3; Gaps 1;

QY 88 acgtgtgtccccgcagcgggagacgcccaccccttgcttctgttgatgtcacccggtccc 147
DB 1123 AGGCTGTGGACATACACCCACAGCTTCCAGCCTGAGCTTGTCTGACACACACAGGCAGT 1182
QY 148 atgttgagcgtctgtatgcagtgatgcagcgcgcctcacgattcttgagcgcagctgtg 207
DB 1183 ATGGCGAGGAATCAACGACGCCAAGATCCAGGCTGCCCGCATTTGTGGAGCAGCGTCAA 1242
QY 208 agcagccgcagccgggttcacgccaactatgcgtgtgtgcctttccacgacccagacatt 267
DB 1243 GGCAGCCCCAT---GGAGCCTGTTTCTATATCTAGTGCCCTTCCAGACCCAGGGGTTT 1299
QY 268 gcccagtgaccctcagcggcgagaccagtggtgtttcagagagagcgtgagacaactctat 327
DB 1300 GCCTCCGCTTTTACAAACAGCGACCCGACAGCTTTTGGCAGAAACTCAACGAGATCCAT 1359
```

```
QY 328 gttcaggaggtgtgtgactgccccagaaatgagtggtggggccatcaaggctccggtggag 387
DB 1360 GCCTTGGGGGTGGAGATGACGACGAGAGATGTGCTGTCTGCTGAGCTAGACCCCTGTTG 1419
QY 388 gttgcacaccccggtctcttcacgtctcttcacgtctcttcacgtctcttcacgtctcttc 447
DB 1420 CACACCCCTCCCTCTCTGACATCTTTGTCTTCTCAGTCAGCTCACCACGAAGATGCTCTT 1479
QY 448 aagaagaatgagctctctgcagctccctgcagctgaagcagtcgagtggtcttcgtgctg 507
DB 1480 CTTACCAACGGGTGGAATCCCTGACTCGGAGAGAGCGCTGCAGGCTGACATTTCTAGTA 1539
QY 508 actggggag 516
DB 1540 ACTGAAGAC 1548
```

```
RESULT 11
AL135796
LOCUS Human DNA sequence from clone GSI-164L12 on chromosome 1 Contains
an STS and GSSs, complete sequence.
ACCESSION AL135796
VERSION AL135796.6 GI:7706886
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 157029)
AUTHORS Donnelly,S.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On May 4, 2000 this sequence version replaced gi:7634143.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TRMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
GSI-164L12 is from the library Genome_Systems_ReleaseI
phelobAC11
This sequence is the entire insert of clone GSI-164L12 The true
left end of clone GSI-15303 is at 103097 in this sequence.
```

```
FEATURES
    source
        Location/Qualifiers
            1..157029
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="1"
                /clone="GSI-164L12"
                /clone_lib="Genome_Systems_ReleaseI"
                1..798
                    /note="L1P repeat: matches 1..791 of consensus"
                    1064..1512
                        /note="LINE2 repeat: matches 5602..6150 of consensus"
            repeat_region
            repeat_region
```

```
misc_feature      complement(2521..2710)
/note="match: GSS: Em:AQ761942"
2792..3250
/note="match: GSS: Em:AQ246643"
3434..3489
/note="14 copies 4 mer tgt 87% conserved"
3435..3488
/note="27 copies 2 mer tg 88% conserved"
4119..4244
/note="MIR repeat: matches 23..152 of consensus"
6902..7098
/note="MER20 repeat: matches 8..218 of consensus"
7164..7191
/note="7 copies 4 mer ttat 96% conserved"
7500..7576
/note="L2 repeat: matches 2645..2739 of consensus"
7735..8041
/note="AluSq repeat: matches 1..308 of consensus"
9440..9901
/note="MLT1D repeat: matches 1..505 of consensus"
10595..10771
/note="MLT2 repeat: matches 277..553 of consensus"
10808..11111
/note="MLT2 repeat: matches 1..289 of consensus"
12089..12200
/note="L2 repeat: matches 2607..2749 of consensus"
12255..12361
/note="LIPB3 repeat: matches 6044..6150 of consensus"
12362..12675
/note="MER1B repeat: matches 1..337 of consensus"
12676..12916
/note="LIPB3 repeat: matches 5804..6044 of consensus"
13006..13612
/note="LIM1 repeat: matches 5180..5793 of consensus"
13614..13877
/note="AluJo repeat: matches 27..287 of consensus"
13879..13981
/note="LIM2 repeat: matches 2048..2147 of consensus"
13982..14292
/note="AluA5 repeat: matches 1..311 of consensus"
14293..15642
/note="LIM2 repeat: matches 640..2048 of consensus"
15608..15818
/note="LIPAL5-16 repeat: matches 146..337 of consensus"
15828..16373
/note="LIPBA repeat: matches -752..-208 of consensus"
16464..16704
/note="LIM2 repeat: matches -676..-428 of consensus"
16497..17042
/note="match: GSS: Em:AQ004976"
17721..19291
/note="L2 repeat: matches 1051..2708 of consensus"
19471..19528
/note="29 copies 2 mer aa 81% conserved"
complement(21071..21467)
21985..22024
/note="match: GSS: Em:AQ761948"
22384..22611
/note="20 copies 2 mer ct 77% conserved"
22697..22863
/note="MIR repeat: matches 3..262 of consensus"
23754..23861
/note="AluJb repeat: matches 137..302 of consensus"
24319..24381
/note="27 copies 4 mer acat 61% conserved"
24382..24670
/note="AluJb repeat: matches 84..148 of consensus"
24671..24822
/note="AluSx repeat: matches 22..309 of consensus"
26950..26985
/note="AluJb repeat: matches 148..289 of consensus"
26952..26987
/note="18 copies 2 mer tg 97% conserved"

repeat_region    /note="9 copies 4 mer tgt 97% conserved"
27687..27917
/note="AluJo repeat: matches 83..304 of consensus"
29345..29476
/note="L2 repeat: matches 2605..2750 of consensus"
30393..30828
/note="match: GSS: Em:AQ825928"
30990..31275
/note="AluSx repeat: matches 14..299 of consensus"
32321..32866
/note="match: GSS: Em:AQ418559"
33201..33344
/note="L2 repeat: matches 2562..2696 of consensus"
33363..33795
/note="L2 repeat: matches 2250..2690 of consensus"
35271..35548
/note="AluSq repeat: matches 18..297 of consensus"
complement(37223..37713)
/note="match: GSS: Em:AQ667379"
37319..37339
/note="match: STS: Em:G37800"
38212..38679
/note="LIMC1 repeat: matches 5847..6324 of consensus"
40470..40593
/note="MIR repeat: matches 8..140 of consensus"
41014..41123
/note="MIR repeat: matches 20..141 of consensus"
41395..41618
/note="MIR repeat: matches 10..256 of consensus"
43230..43494
/note="AluY repeat: matches 39..307 of consensus"
43683..44142
/note="LIMEC repeat: matches 1211..1677 of consensus"
44767..45295
/note="L1 repeat: matches 2530..3129 of consensus"
45533..45571
/note="AluJb repeat: matches 3..41 of consensus"
45604..46929
/note="LIPA7 repeat: matches 4758..6124 of consensus"
46930..47199
/note="AluJb repeat: matches 24..297 of consensus"
47358..47823
/note="L1 repeat: matches 3647..4127 of consensus"
47918..48310
/note="L1 repeat: matches 4257..4654 of consensus"
48311..48614
/note="AluSx repeat: matches 1..308 of consensus"
48615..48796
/note="L1 repeat: matches 4654..4834 of consensus"
49029..49516
/note="match: GSS: Em:AQ280129"
49087..49262
/note="LIP repeat: matches 5164..5357 of consensus"
49407..49510
/note="L2 repeat: matches 2579..2696 of consensus"
50481..51060
/note="MLT1F repeat: matches 1..536 of consensus"
complement(51651..52071)
/note="match: GSS: Em:AQ636105"
52290..52512
/note="match: GSS: Em:AQ797420"
52936..53296
/note="LIPAL5 repeat: matches 5785..6156 of consensus"
53303..53330
/note="7 copies 4 mer tgt 96% conserved"
53355..53457
/note="LIPAL5 repeat: matches 5689..5791 of consensus"
53680..53895
/note="108 copies 2 mer tt 57% conserved"
complement(54152..54622)
/note="match: GSS: Em:AQ568896"
54157..54274
/note="MER45B repeat: matches 199..318 of consensus"
```



```

FEATURES
  source
    1. 717
      Location/Qualifiers
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="9"
        /sex="male"
        /clone="SC9-12pJ34g10"
        /tissue_type="Lymphoblastoid cell line"
        /clone_lib="SC9-12pJ"
      186 a 198 c 211 g 120 t 2 others
      ORIGIN

Query Match          7.98; Score 65.4; DB 11; Length 717;
Best Local Similarity 72.4; Pred. No. 0.0036;
Matches 84; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 235 tatcgctgggtgccttttcacgacccagacattggcccgtagcctcacgcggagacca 294
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 557 TCTGGCGTCCCTCTCTCTCTCTCCACAGATATTGGCNCAGTGACCTTCACGGGGACCCC 498

Qy 295 gtggtgttcagagagcgtgagacaaactctatgttcaggagggtggtgactgcc 350
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 497 ACAGTGTGTTTCAGAGGAGCTCAGAGAACTTACGTGTCAGGTGGGCAGCGCCTTGACC 442

```

Search completed: March 7, 2002, 20:47:59
Job time: 7212 sec

[illegible]

Search completed: March 7, 2002, 20:47:59
Job time: 7212 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 21:17:38 ; Search time 148.15 Seconds
(without alignments)
4774.170 Million cell updates/sec

Title: US-09-665-728-2_COPY_96_920

Perfect score: 825

Sequence: 1 atgacgcctaggcgagcgtct.....ctcctcttctgatgcaagac 825

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 42862619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_1101.*
- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	825	100.0	1031	22	AAF82464
2	51.6	6.3	2004	18	AAF82464
3	50.4	6.1	2338	19	AAF82464
4	50.4	6.1	2338	21	AAF82464
5	47.2	5.7	2338	12	AAF82464
6	43.4	5.3	2745	22	AAF82464
7	41.8	5.1	9063	22	AAF82464
8	41.2	5.0	1260	19	AAF82464
9	40.8	4.9	1160	20	AAF82464
10	40.4	4.9	2271	15	AAF82464
11	40.4	4.9	2271	15	AAF82464

12	40.2	4.9	2214	19	AAF82464
13	40.2	4.9	3331	19	AAF82464
14	40.2	4.9	3331	20	AAF82464
15	40.2	4.9	3331	22	AAF82464
16	39.6	4.8	2502	16	AAF82464
17	39.6	4.8	2502	19	AAF82464
18	39.6	4.8	2502	20	AAF82464
19	39.6	4.8	4748	22	AAF82464
20	38.8	4.7	1929	19	AAF82464
21	38.8	4.7	2668	21	AAF82464
22	38.6	4.7	1598	14	AAF82464
23	38.2	4.6	1995	12	AAF82464
24	38.2	4.6	1995	19	AAF82464
25	38.2	4.6	1995	21	AAF82464
26	38	4.6	1588	22	AAF82464
27	37.8	4.6	1747	20	AAF82464
28	37.6	4.6	567	21	AAF82464
29	37.4	4.5	571	20	AAF82464
30	37.4	4.5	3120	21	AAF82464
31	37.4	4.5	5145	21	AAF82464
32	37.4	4.5	6085	18	AAF82464
33	37.4	4.5	31422	21	AAF82464
34	37.2	4.5	355	21	AAF82464
35	37.2	4.5	5335	21	AAF82464
36	37	4.5	2000	18	AAF82464
37	37	4.5	42000	21	AAF82464
38	37	4.5	63164	21	AAF82464
39	36.8	4.5	1886	20	AAF82464
40	36.8	4.5	1886	22	AAF82464
41	36.8	4.5	24379	18	AAF82464
42	36.8	4.5	24379	19	AAF82464
43	36.8	4.5	50937	21	AAF82464
44	36.6	4.4	426	7	AAF82464
45	36.6	4.4	718	22	AAF82464

ALIGNMENTS

RESULT 1
AAF82464
ID AAF82464 standard; cDNA; 1031 BP.
XX
AC AAF82464;
XX
DT 29-JUN-2001 (first entry)
XX
DE Rat cDNA clone P00210D09.
XX
KW Rat; secreted factor; P00210D09; cardiant; nephrotropic;
KW antiinflammatory; gene therapy; cardiac disease; renal disease;
KW inflammatory disease; ss.
XX
OS Rattus norvegicus.
XX
FH Key Location/Qualifiers
FT CDS 96..923
FT /*tag= a
FT /*product= "Rat secreted factor"
FT sig_peptide 96..158
FT /*tag= b
FT mat_peptide 159..923
FT /*tag= c
WO200123419-A2.
05-APR-2001.
27-SEP-2000; 2000WO-US26582.
27-SEP-1999; 99US-0156277.
(SCIO-) SCIOS INC.

XX Example 3: Fig 6A-D; 65pp; English.
 XX The invention provides isolated cDNA molecules coding for spider silk
 CC proteins. The spider silk proteins are characterized by repeating alpha
 CC and beta regions and optional variable regions. The DNA sequences are
 CC useful in the production of spider silk protein by recombinant DNA
 CC techniques. The recombinant spider silk proteins may be used for the
 CC production of fibers. The present sequence represents the cDNA encoding
 CC the spider silk protein 1, derived from the major ampullate gland of
 CC Nephila clavipes.
 XX
 XX Sequence 2338 BP; 564 A; 432 C; 918 G; 424 T; 0 other;

Query Match 6.1%; Score 50.4; DB 21; Length 2338;
 Best Local Similarity 49.3%; Pred. No. 0.0088;
 Matches 132; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
 Qy 419 tctcggatgcccgtgccaaagactaccacaaagaagaatgagctcctcgagctcctgcagc 478
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 459 tcttggaaagcgaagtgccagcagaggtggattaggtggacaaaggtgcagtgccagc 518
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 479 tgaagcagtcagagtggtctgtcgtgactggggagctgcggtgacccgacccaccctg 538
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 519 agcagcagccggaggtgctggacaaagcggtacgtggtgttctgtggacaaaggtgccg 578
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 539 gctacctggtctttgagagatgcctccaccagttcttggcacaaggtgtccagctgggaca 598
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 579 acaaggaggtcataggagacttgaagcacaaggtgctggacagagagattagtgaggaca 638
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 599 agcagcaggtgctggaggtgttaaagtgggtggagtcgcgcacccagcctccaaagttc 658
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 639 aggtgcaggtgcagcagcagcagcagctggaggtgcccggacaaaggagactaggtgg 698
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 659 atctgctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 686
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 699 acaaggtgctggacaaaggagctggagca 726

RESULT 5

AAQ14183
 ID AAQ14183 standard; cDNA; 2338 BP.

AC AAQ14183;

15-JAN-1992 (first entry)

N.clavipes dragline silk protein coding sequence.

protein superfbre; major ampullate silk; orb web spider; ss.

Nephila clavipes.

Key Location/Qualifiers
 CDS 1..2157
 /*tag= a
 /product= silk protein 1

EP452925-A.

23-OCT-1991.

18-APR-1991; 91EP-0106217.

20-APR-1990; 90US-0511792.

(UWY-) UNIV OF WYOMING.

Lewis RV, Xu M, Hinman M;

WPI; 1991-312199/43.

P-PSDB; AAR14308.

DR

XX DNA encoding spider silk protein-1 and 2 and variants - isolated
 PT from Nephila clavipes, for prodn. of spider silk protein and
 PT fibres having desired characteristics
 XX
 PS Claim 4; Page 23; 48pp; English.
 XX
 CC A N.clavipes major ampullate gland cDNA library was screened with
 CC probes based on peptide fragments of the purified spider silk
 CC protein (see AAQ14185). Positive plaques were identified and the
 CC spider silk protein 1 coding sequence was determined. See also
 CC AAQ14184.
 XX
 XX Sequence 2338 BP; 566 A; 433 C; 916 G; 423 T; 0 other;

Query Match 5.7%; Score 47.2; DB 12; Length 2338;
 Best Local Similarity 48.5%; Pred. No. 0.051;
 Matches 130; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
 Qy 419 tctcggatgcccgtgccaaagactaccacaaagaagaatgagctcctcgagctcctgcagc 478
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 459 tcttgaagcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 518
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 479 tgaagcagtcgaggtggtcttctgctgactggggagctgcggtgacccgacccaccctg 538
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 519 agcagcagccggaggtgctggacaaagcggtacgtggtgttctgtggacaaaggtgccg 578
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 539 gctacctggtctttgagagatgcctccaccagttcttggcacaaggtgtccagctgggaca 598
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 579 acaaggaggtcataggagacttgaagcacaaggtgctggacagagagattagtgaggaca 638
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 599 agcagcaggtgctggaggtgttaaagtgggtggagtcgcgcacccagcctccaaagttc 658
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 639 aggtgcaggtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 686
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 659 atctgctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 686
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 699 acaaggtgctggacaaaggagctggagca 726

RESULT 6

AAQ09568

ID AAD09568 standard; cDNA; 2745 BP.

AC AAD09568;

10-SEP-2001 (first entry)

Human transporter and ion channel-17 (TRICH-17) cDNA.

Human; transporter and ion channel-17; TRICH-17; cystic fibrosis; mood;
 gene therapy; amyotrophic lateral sclerosis; amnesia; muscular dystrophy;
 hypertension; angina; neurological disorder; asthma; bipolar disorder;
 dementia; depression; Alzheimer's disease; epilepsy; vaccine; arrhythmia;
 Pick's disease; ischaemic cerebrovascular disease; AIDS; anxiety; stroke;
 Huntington's disease; Parkinson's disease; cerebral neoplasm; allergy;
 demyelinating disease; mental disorder; Schizophrenia; polymyositis;
 muscle disorder; cardiomyopathy; cataract; myocarditis; Grave's disease;
 dermatomyositis; diabetes mellitus; immunological disorder; psoriasis;
 rheumatoid arthritis; Sjogren's syndrome; systemic lupus erythematosus;
 sickle cell anaemia; Wilson's disease; infertility; Cushing's disease;
 scleroderma; pulmonary artery stenosis; neutropenic; Addison's disease;
 malabsorption syndrome; hypercholesterolaemia; cancer; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 246..2066

/*tag= a

/product= "Human TRICH-17 protein"

WO200146258-A2.

Qy 490 caggtggtcttcgtgactgggagtg 518
Db 8063 gccctgaacgcgcgtgcagagtgagtg 8091
RESULT 8
AAV41730
ID AAV41730 standard; DNA; 1260 BP.
XX
XX
AC AAV41730;
XX
XX
DT 20-NOV-1998 (first entry)
XX
DE Codon-optimised RAmY3D signal fused to DNA encoding mature AAT.
XX
KW Protein expression; monocotyledon plant cell;
KW glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;
KW AATIII; human serum albumin; HSA; subtilisin BPN'; treatment; emphysema;
KW antithrombotic; blood replacement; ss.
XX
XX Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1..75 /*tag= a
FT /*note= "codon-optimised RAmY3D signal sequence"
FT misc_feature 76..1260 /*tag= b
FT /*note= "encodes mature AAT"
XX
XX WO9836085-A1.
XX
XX PD 20-AUG-1998.
XX
XX PF 13-FEB-1998; 98WO-US03068.
XX
XX PR 13-FEB-1997; 97US-00381170.
XX PR 13-FEB-1997; 97US-0037991.
XX PR 13-FEB-1997; 97US-0038169.
XX PR 13-FEB-1997; 97US-0038169.
XX
XX (PHYT-) APPLIED PHYTOLOGICS INC.
XX
XX Rodriguez RL, Sutliff TD;
XX
XX WPI; 1998-467179/40.
XX
XX Expressing mature, glycosylated proteins in monocotyledonous plant
XX cells - from chimeric gene including signal peptide sequence,
XX specifically therapeutic agents and industrial enzymes
XX
XX Disclosure; Pages 34 iii-iv; 53pp; English.
XX
XX The present sequence encodes a fusion protein of codon-optimised RAmY3D
XX signal sequence/mature alpha 1-antitrypsin (AAT). The protein is used
XX to exemplify the invention. The specification describes a method for
XX producing mature heterologous protein in monocotyledonous plant cells.
XX The method comprises transforming the cells with a chimeric gene
XX comprising a monocotyledon transcription regulator, inducible either
XX during seed maturation or by adding/removing a small molecule, DNA
XX encoding the heterologous protein, and DNA encoding a signal peptide,
XX with the signal peptide causing secretion of the protein from the cell.
XX Proteins expressed in this manner include mature glycosylated alpha
XX 1-antitrypsin (AAT) with a glycosylation pattern that significantly
XX increases its serum half-life, mature glycosylated antithrombin III
XX (AATIII), mature human serum albumin (HSA) having the native folding
XX pattern as shown by bilirubin-binding characteristics, or mature active
XX subtilisin BPN'. These proteins are useful therapeutically (e.g. AAT for
XX treating emphysema, AATIII as antithrombotic and HSA as blood replacement)
XX or as industrial enzymes (BPN' is used in detergents).

SO Sequence 1260 BP; 287 A; 428 C; 350 G; 195 T; 0 other;
Query Match 5.0%; Score 41.2; DB 19; Length 1260;
Best Local Similarity 47.6%; Pred. No. 1.2;
Matches 121; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
Qy 369 catacaggtgcgcgtggaggtgtccaaaccccggtctctcatctatcgtctctcgatgc 428
Db 759 catacagcactgcaagaagctctccagctgggtgctctcatgaagctacctggggaacgc 818
Qy 429 ccgtgccaaagactaccacaagaagaatgagctcctcagctcagctcagctgaacagtc 488
Db 819 caccgcacattctctcctgcgcagagggcaagctccacacctggagaaacgagctgac 878
Qy 489 gcaggtggtcttcgtgctgactgggactcgggtgaccgcacccacctggtacctggc 548
Db 879 gcacgacatcatcagcaagttctctggagaacgaggaagcgctcctagctccacct 938
Qy 549 ttttgaggagatgcctccaccagttcttgcccaagtgttccagctggacaagcaggt 608
Db 939 ccggaagctgagcatcacgcgcagctacgctgacgctgaagcgtgctggccagctggc 998
Qy 609 gtcggaggtgttaa 622
Db 999 cacgaaggtcttca 1012
RESULT 9
AAV25208
ID AAV25208 standard; CDNA; 1160 BP.
XX
XX AC AAV25208;
XX
XX DT 19-JUL-1999 (first entry)
XX
XX DE Maize caffeoyl-CoA 3-O-methyltransferase cDNA.
XX
XX KW Maize; corn; caffeoyl-CoA 3-O-methyltransferase; lignin;
XX KW transgenic plant; ss.
XX
XX OS Zea mays.
FH Key Location/Qualifiers
FT CDS 167..943 /*tag= a
XX
XX PN WO9910498-A2.
XX
XX PD 04-MAR-1999.
XX
XX PF 24-AUG-1998; 98WO-US17519.
XX
XX PR 12-MAY-1998; 98US-0076851.
XX PR 27-AUG-1997; 97US-0057082.
XX
XX (PTON-) PIONEER HI-BRED INT INC.
XX
XX Bowen BA, Helentjaris TG, Wang X;
XX WPI; 1999-204667/17.
XX P-FSD; AAY05669.
XX
XX Nucleic acids encoding plant lignin biosynthesis enzymes - used to
XX transform plants to modulate lignin biosynthesis
XX
XX Claim 2; Page 138-139; 166pp; English.
XX
XX This DNA sequence encodes a caffeoyl-CoA 3-O-methyltransferase (see
XX AAY05669) of maize. A polynucleotide having this sequence can be
XX amplified from a cDNA library prepared from shoot culture of
XX maize line CM45 using the primers given in AAX25241 and AAX25242. The
XX invention provides methods and compositions relating to altering

DR P-PSDB; AAW56163.

XX New cDNA and e.g. vector, host cell and polypeptide - used to

PT produce polypeptide in high yields, which is used in cosmetics

XX Claim 2; Pages 7-9; 15pp; Japanese.

XX The present sequence represents a new DNA sequence isolated from Pinctada

CC fucata. The encoded polypeptide be used as an ingredient in cosmetics.

XX Sequence 3331 BP; 762 A; 701 C; 1022 G; 846 T; 0 other;

XX

Query Match 4.9%; Score 40.2; DB 19; Length 3331;

Best Local Similarity 51.4%; Pred. No. 2.5;

Matches 93; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 458 agctctgcagctcctgcagtcgaagcagtcgaggtggtcttctgctgactgggact 517

Db 766 agccgcagcgcgtctgccgtgcaggagcgcaggtggaacttgaggactcgtgccc 825

Qy 518 gcggtgaccgcacccaccctggtacctgcttttgaggagatcgctccaccagttctg 577

Db 826 acttgaggactcgtgagcacttgaggcctcgaggtcttggctcgcctcgaggata 885

Qy 578 gccagtgctccagtcgacacagcagcaggtgctcgaggtgttaagtgggtgagtcgc 637

Db 886 tggaggatctgctgccgtcgtctgctgctgccgcgctgctgccggaggaggact 945

Qy 638 c 638

Db 946 c 946

RESULT 14

AAZ32021

ID AAZ32021 standard; DNA; 3331 BP.

AC AAZ32021;

XX 10-JAN-2000 (first entry)

XX Human METH1 related EST D86074.

DE Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;

XX cancer; diagnosis; hyperproliferative disorder; autoimmune disease;

XX angiogenesis inhibitor; abnormal wound healing; inflammation;

XX rheumatoid arthritis; psoriasis; endometrial bleeding disorder;

XX diabetic retinopathy; macula degeneration; haemangioma; detection;

XX arterial-venous malformation; immune deficiency; ss.

OS Homo sapiens.

XX WO9937660-A1.

XX 29-JUL-1999.

XX 22-JAN-1999; 99WO-US01313.

XX 23-JAN-1998; 98US-0072298.

XX 28-AUG-1998; 98US-0098539.

XX (IRUE/) IRUELA-ARISPE L.

XX (HAST/) HASTINGS G A.

XX (RUBE/) RUBEN S M.

XX IrueLA-Arispe L, Hastings GA, Ruben SM;

XX WPI; 1999-590684/50.

XX New isolated metalloprotease thrombospondin polypeptides, useful for

PT treating hyperproliferative disorders, cancers or autoimmune disorders

PT

XX Disclosure; Page 322-324; 457pp; English.

XX AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human

CC metalloprotease thrombospondin (METH) proteins METH1 and METH2

CC respectively. METH1 and METH2 have been found to be potent inhibitors of

CC angiogenesis both in vitro and in vivo. They can be used for treating

CC cancer and other disorders related to angiogenesis including abnormal

CC wound healing, inflammation, rheumatoid arthritis, psoriasis,

CC endometrial bleeding disorders, diabetic retinopathy, some forms of

CC macula degeneration, haemangiomas, and arterial-venous malformations.

CC They may be useful in treating deficiencies or disorders of the immune

CC system, by activating or inhibiting the proliferation, differentiation,

CC or mobilisation (chemotaxis) of immune cells. The etiology of these

CC immune deficiencies or disorders may be genetic, somatic, such as

CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or

CC toxins), or infectious. They can also be used to treat inflammatory

CC conditions, both chronic and acute conditions. The products can also be

CC used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to

CC AAY49511 represent sequences given in the exemplification of the present

CC invention.

XX Sequence 3331 BP; 762 A; 701 C; 1022 G; 846 T; 0 other;

XX

Query Match 4.9%; Score 40.2; DB 20; Length 3331;

Best Local Similarity 51.4%; Pred. No. 2.5;

Matches 93; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 458 agctctgcagctcctgcagtcgaagcagtcgaggtggtcttctgctgactgggact 517

Db 766 agccgcagcgcgtctgccgtgcaggagcgcaggtggaacttgaggactcgtgccc 825

Qy 518 gcggtgaccgcacccaccctggtacctgcttttgaggagatcgctccaccagttctg 577

Db 826 acttgaggactcgtgagcacttgaggcctcgaggtcttggctcgcctcgaggata 885

Qy 578 gccagtgctccagtcgacacagcagcaggtgctcgaggtgttaagtgggtgagtcgc 637

Db 886 tggaggatctgctgccgtcgtctgctgctgccgcgctgctgccggaggaggact 945

Qy 638 c 638

Db 946 c 946

RESULT 15

AAC90078

ID AAC90078 standard; DNA; 3331 BP.

XX AAC90078;

XX 19-MAR-2001 (first entry)

XX D86074 cDNA clone.

XX METH; metalloprotease; thrombospondin; angiogenesis inhibition;

XX cancer therapy; benign tumour; ocular angiogenic disease;

XX rheumatoid arthritis; psoriasis; wound healing; endometriosis;

XX vasculogenesis; granulation; hypertrophic scar; nonunion fracture;

XX scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;

XX coronary collateral; cerebral collateral; arteriovenous malformation;

XX ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;

XX plaque neovascularisation; telangiectasia; haemophilic joint; EST;

XX angiofibroma; fibromuscular dysplasia; expressed sequence tag;

XX Crohn's disease; atherosclerosis; birth control; ss.

OS Unidentified.

XX WO200071577-A1.

XX 30-NOV-2000.

XX

